

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner #: _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

	Type of Search	Vendors and cost where applicable
Searcher: <u>1664 Port</u>	NA Sequence (#) <u>1</u>	STN _____
Searcher Phone #: <u>300-3537</u>	AA Sequence (#) <u>1</u>	Dialog _____
Searcher Location: _____	Structure (#) _____	Questel/Orbit _____
Date Searcher Picked Up: <u>4/11</u>	Bibliographic _____	Dr. Link _____
Date Completed: <u>4/13</u>	Litigation _____	Lexis/Nexis _____
Searcher Prep & Review Time: <u>15</u>	Fulltext _____	Sequence Systems <u>gs</u>
Clerical Prep Time: _____	Patent Family _____	WWW/Internet _____
Online Time: <u>10</u>	Other _____	Other (specify) _____

THIS PAGE BLANK (USPTO)

```

RESULT 1
T34105
hypothetical protein C17G10.8 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34105
R:Johnson, D.
submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid C17G10.
A:Reference number: Z21476
A:Accession: T34105
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-938 <J0H>
A:Cross-references: EMBL:U028739; PUDN:AAB93456.1; GSPDB:GMD00020; CESP:C17G10.8
A:Experimental source: strain Bristol N2; clone C17G10
C:Genetics:
A:Gene: CESP:C17G10.8
A:Map position: 2
A:Introns: 25/1; 67/2; 158/3; 214/3; 329/3; 386/2; 465/3; 510/1; 525/2; 582/3; 709/1

Query Match          48.3%, Score 1034, DB 2; Length 938;
Best Local Similarity 54.6%, Pred. No. 5e-68;
Matches 230; Conservative 52; Mismatches 111; Indels 28; Gaps 8;

QY      7  RLACGTATTTGASRIGKATLAKAKDGNANIYIAKTKNQPHKLLGTITYTAEELEAVNG 66
Db      526  KFGVGTIVLTITGSRIGKEIKELAKLAKDGNANIYAAKTKTAHKKPGTITSAAEETEKAGG 585

QY      67  KALPCTIVDVDRBOOQSAAVEAKAIKKFGGIDILVNNASAISSLTNTLDTPETKRLDLMANNYT 126
Db      586  KALPCTIVDVDRBASVAKAAVEAEVAKKFGGIDILINNASAISSLTNTENTEMKRIIDLHNSINT 645

QY      127  KGTYLAKACIPYLYKSKSVAHILNTISPLLNLNPNWFKOHCAYTIAKYGMSMYVLGMAEEF 186
Db      646  RGTFLMTKTCPLPYLKSGRNPHVLTISPMLMETRFWFAHVAVYTAAYKYSKMCVLQGBEEF 705

QY      187  KGE-IAVAALMPKTAIRHAAADMGLGPGIESOCRRVYDIADAANSIFOK-PRSLFGNRYI 244
Db      706  RPHGIAVAALMLPLTAIWWAAEMLSDKGEGESRRKPSIAAQAAYVLTAKNSNDFDFTGNCCI 765

QY      245  DENILKEGSIENPDYVAIKPGHLPDPDFLEDPYPAVSKVSTGAVPEFKREKIQLOPK 304
Db      766  DEDILKAGVTDFTRIACVDPAPLMPDFFI---PAGTYDHKFFSSA-----QIGKK 813

QY      305  PR---SGAVEE---TFRIIVKDSLSDDVAVKATQALYIELLSGEDGCT-----WFLDKSK 352
Db      814  NKTGAGVAVEEIKQIFLSAKRLNLADIVYKTKGFYEFL--KDPPTKSERITITDLKNG 871

QY      353  GGNNGYGEPSDQADVYMSKMTDDPYKMFSGKILPMAVMSGKILTKGMALAIKIEKLMN 412

```

Db	872	EGALTDKNA	SKADYKFTL	AEHFAPLET	GKLRPTTAL	MTKKQISD	MPGAMKLES	LR	931
QY	413	Q	413						
Db	932	K	932						

RESULT 2
T119954
hypothetical protein C45B11.3 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T119954
R:McMurray, A.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19202
A:Accession: T119954
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-293 <WTL>
A:Cross-references: EMBL:Z74029; PIDN:CAA98431.1; GSPDB:GN00023; CESP:C45B11.3
A:Experimental source: clone C45B11
C:Genetics:
A:Gene: CESP:C45B11.3
A:Map position: 5
A:Introns: 9/2: 66/3
A:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Query Match	41.78;	Score 891.5;	DB 2;	Length 293;
Best Local Similarity	61.78;	Pred. No. 3e-58;		
Matches 185;	Conservative 29;	Mismatches 73;	Indels 13;	Gaps 3;

Qy	2	LPNPGRLGCGVFTFTGASRSGIGKAIATAKAADGANIYAAATAPRHKLGITITAAEEI	61
		:::	
Db	5	IKNNGFPAKGKVFITIGASRSGIGKEIATAKADGANIYAAKATAHRHKLGITITAAEI	64
Qy	62	EAVGGKALPCIVDYRDEQOISAAVEKAIRKKGIDILYNNASATSLNTLDTPRKLDLM	121
Db	65	EKAGSHALPCVADVDRDEAAVAAADAAPKKGEGDILINNASATSLNTEGTDMRKYDLM	124
Qy	122	MNVNTRGTYLYASKRCIPPLKSKVAHLITNSPRLNLPVMEKOHCAVTAIKYCSMYLG	181
		:::	
Db	125	HSINTRGTYLLTKCLPLKKGKMPHYNISPRIDMEAKMGFPVGTAKMAFGSMCYLG	184
Qy	182	MAEEFGK-EIAVNALMPKTAIHTAMDMLGGPGIESOCRKVDIIDAAYAS::IFOK-PRKFT	239
		:::	
Db	185	HHEEFPRPGIAYNALMPPLTAITWTSAMEPLSHGSDSAGNRKASIMADSAVALISDSKKT	244
Qy	240	GNFVDEIMLEKEBGIENDVYAIRPGHLPLODFLDEYPRKAVSKKVESTAVPRFKREKL	299
		:::	
Db	245	GNFLDEELLDTGISTITNEQECYVGGSSLIPDFEVPKSFV-----PFGGSKL	293

RESULT 3
G83284
Probable short-chain dehydrogenase PA2892 [Imported] - *Pseudomonas aeruginosa* (strain PA
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence:revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83284
R:Stover, C.K.; Plam, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Laribig, K.; Llm,
L.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: G83284
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-274 <SMO>
A:Cross-references: GB:AE004715; GB:AE004091; NID:99948977; PIDN:AMG06280.1; GSPDB:GN001
A:Experimental source: strain PA01
A:Genetics:

Query Match 33.18; Score 708; DB 2; Length 274;
 Best Local Similarity 54.08; Pred. No. 8.5e-45;
 Matches 147; Conservative 44; Mismatches 77; Indels 4; Gaps 4;
 A:Gene: PA2892 ribitol dehydrogenase; short-chain alcohol dehydrogenase homology
 C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

OY	8	LACGCVFTTGSRSRGKATLKAADGANIYLAATQAPHPKLGTITYAAEEIEVAGK	67
Dd	3	LHGKLFITGASRSGRGREIALRAARDGANILYAASAEPRHKBESTIISVAEYVMAQG	62
OY	68	ALPCIVDVDEQOISAAVEKAIKKEFGIDILVNNAISLTNTLDTPPKRLDLMNVNR	127
Dd	63	ALPILDVRDEQAVAAAARAEFEGIDALVNNGAIRLVGEKLEBKRPFDLMQINTR	122
OY	128	GTYLASKRCIPPLKSKVAHLINTISSPILNLPVMVEKHCHATTIAKGYSMVLGAEEF	186
OY	123	AVLVSOQAALPLRKNSANGHILTSPPINTLGWRMHOQPPTVKYGNMSLLTGHEEFG	182
OY	187	KGEIVNALMPKTAIHTAAMDMLGGPGIESOCRRVDIADANYSIF-QKPSPFTGNFI	244
Dd	183	KYAI SVNALMPMTATATAAIEFELCSRDFAFRARPATIMADAHAHILTSSEGRSLGLLV	242
OY	245	DENILKEGIEMEDYALKP-CHPLQDPDFLD	275
Dd	243	DEELLERKQSDEFQRYRDPREGALVPDLFLD	274

4
RESULT
A:70597
hypothetical protein RV3224 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: A70597
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
J.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: A70597
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-782 <COL>
A:Cross-references: GB:295120; GB:AL123456; NID:q3361739; PIDN:CAB08313.1; PID:q20726
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: RV3224
C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

	Query Match	31.1%; Score 665; Db 2;	length 282;
	Best Local Similarity	48.9%;	Pred. No. 1,3e-41;
	Matches	133; Conservative	57; Mismatches 78; Indels 4; Gaps
Oy	8 LAGCTVETTGASRGIGAIKAAKDGANIVIAKTQAPHPKLITGYTAABEIEAVGK	67	
Dd	3 LGNKTMFISGSRCIGLILARRARDGNAILIKTEBPBKLPCTVFATAKEELENGO	62	
Oy	68 ALPCTVDVRDGOISAANEKAIRKKFGGDILVNNAASLSLTNLDPTFKRLDIMMNVNR	127	
Dd	63 ALPIYGDTRDDPAVASAATVTEDQGGLDICNNASNAINSGISTEVPMKRFIDLMNGIQVR	122	
Oy	128 GTYLASKACIPELYLKRSKVAYHLNISPPRLNPVPNFKHCAVTIAKYGMSMYLVGAEEFK	187	
Dd	123 GTYAASQCACIPHHMKGREPNHILTSLPILLEKKWLRP-TAYMAAKYGMTLCALGIAEEMR	181	
Oy	188 GE-IANVALMPKRTAHITPAAM-DMLGGECIESCCAKRVIIIDAASIFQRKS-FTGNFVI	244	
Dd	182 ADGIASNTLMPTRWATVAADANLLGGDEADAMARSKRPVEYADAADVINKPKATEYTGTTL	241	

OY 245 DENILKEGIENFDYAIKPGHLPDPFLDE 276
 DB 242 CEDVLVESGVTLSDVDCVPGATLGVLDLWED 273

RESULT 5

716638

hypochemical protein M03A8.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T16638

R:Du, Z.

submitted to the EMBL Data Library, November 1995

A:Description: The sequence of C. elegans cosmid M03A8.

A:Reference number: Z18551

A:Accession: T16638

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-436 <DUZ>

A:Cross-references: EMBL:U41544; NID:g1109886; PID:g1109889; PIDN:AA83184.1; CESP:M03A8

C:Genetics:

A:Gene: CESP:M03A8.1

A:Introns: 92/1; 115/1; 205/3; 361/1; 405/3

Query Match 13.4%; Score 286; DB 2; Length 436;

Best Local Similarity 25.8%; Pred. No. 1.8e-13;

Matches 120; Conservative 73; Mismatches 176; Indels 96; Gaps 17;

OY 7 RLACTGVTGTGASRGIGKALALKAADGANIYAAKTAOPHPKLLCTIYTA-----EE 60
 DB 4 RPDGKVAIVTGAGGGLGKTYALELAKRGCKVYVNDLGDRH---CTSSSSMAKRVGE 59

OY 61 IEAVGKALPCIVDVRDEQOISAAYEKAIRKGGIDILVNNASISLTNTLTPTKRLDL 120
 DB 60 IKSAGQAAVANYDSVEFGDKI---VKTALDNFGRIDIVINNGLRDSVFLKMTLMDL 116

OY 121 MNMVTNRGYLAKACIPYLKKSKVAHIINISPLNLNPMVFKOHCAVITAKGMSMYL 180
 DB 117 IKRVHKGKAVYTKAMPFMRDQKGRIVYVSSNAGVHGNFGQAN--YAAKSAALIGLSN 174

OY 181 GNAEE-FKEGIAVNLMP-----KTAIHTAAMDML----- 209
 DB 175 SLAEGAKNIILANTLVPAAGSRLEFETVPMQNLVDALPKPDYVPLVTVYVHDSFEESGKV 234

OY 210 --GGG-----ISQCKRVIIIDAAVSTIQPKSFNGNVIDENT--LKEEGLENFDVA 261
 DB 235 FEAGAGWYGTIOYKSKGVISHASADDIKAMWSTTTNNNGAEYIGTITEQSARLVSI-- 292

OY 262 IKRPHLPDPFLDEY-----PEAVSKRYESTGAVPEFKEEKLQLOPKRSGAVETFPRI 316
 DB 293 -----LKEHSASSSSSSSSGSSGSGAPPS-----NIRSSAL---FDE 327

OY 317 VKDSLSD--VVKATQATVYLFELSGEDG---GTWFLDLKSKGNGVGE--PSDAQADV 368
 DB 328 MADVKAADPTAVTKLSIYLIIT--DGKNELGKFTLDRKSSPSVYLDVKNKGEKANMT 385

OY 369 MGMTDDPFYKMFSGKIKPTMAFMSGKLKIGKMAALAIKLEKLMNQ 413
 DB 386 VIVADSDVFVIAAGKILNAQAKAFMSGKLKYGKGNVMLQKLTQVLEK 430

RESULT 6
 T44932

3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) [imported] - Agrobacterium tum

C:Species: Agrobacterium tumefaciens

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000

C:Accession: T44932

R:Kim, K.S.; Farrand, S.K.

J. Bacteriol. 178, 3275-3284, 1996

A:Title: T1 plasmid-encoded genes responsible for catabolism of the crown gall opine man

by the plant tumor.

A:Reference number: Z22872; MUID:96236046

A:Accession: T44932

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-248 <KIM>

A:Cross-references: EMBL:U19620; NID:g797330; PIDN:AAB07783.1; PID:g797334

A:Experimental source: strain 15955

C:Genetics:

A:Gene: mocc

A:Genome: plasmid pT115955

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

C:Keywords: oxidoreductase

Query Match 12.5%; Score 267; DB 2; Length 248;

Best Local Similarity 34.6%; Pred. No. 2e-12;

Matches 71; Conservative 41; Mismatches 71; Indels 22; Gaps 8;

OY 7 RLACTGVTGTGASRGIGKALALKAADGANIYAAKTAOPHPKLLCTIYTAEEIYAVG 66
 DB 2 KIQKAVITGAGRGIGRTALELKEGCVYLAIE-----LNEVAVAEVVAIS 54

OY 67 KALPCIVDVRDEQOISAAYEKAIRKGGIDILVNNASISLTNTLTPTKRLDMMNVN 125
 DB 55 EALALRTDVQHKSEVDALAKAEFERGAVDIIVNNA-GVAIHNTIPNIREADMDMMMAIN 113

OY 126 TRGTYASKACIPYLKKSKVAHIINISPLNLNPMVFKOHCAVITAKGMSMYLGMAR- 184
 DB 114 LKGTFLCTAATFQHMCDRRHGHINVVS--RAKVASAFGANASKFGH---LGFOT 167

OY 185 -EFKG-EIAVNLMPKTAIHTAAMD 207
 DB 168 TDQEGIEFQVKA---TAVCPGAVD 188

RESULT 7
 D69930

probable 3-oxoacyl-[acyl-carrier-protein] reductase (EC 1.1.1.100) yoxD - Bacillus su

C:Species: Bacillus subtilis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: D69930; S01270

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brulliet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funai, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scani

A:Authors: Schleich, S.; Schroeder, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

auchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Whitters, P.; Wipet, A.; Yamamoto, H.; Yamane, K.; Yasunoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumein, E.; Yoshikawa, K.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A69580; MUID:98044033

A:Accession: D69930

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-238 <KUN>

A:Cross-references: GB:Z99114; GB:AL009126; NID:g2634230; PIDN:CAB13743.1; PID:g26342

A:Experimental source: strain 168

R:Carigan, C.M.; Haarsma, J.A.; Smith, M.T.; Wake, R.G.

Nucleic Acids Res. 15, 8501-8509, 1987

A:Title: Sequence features of the replication terminus of the Bacillus subtilis chrom

A:Reference number: S01270; MUID:88040469

A:Accession: S01270

A:Molecule type: DNA

A:Residues: 62-238 <CAR>

A:Cross-references: EMBL:X06168; NID:g40205; PIDN:CAA29533.1; PID:g809662

C:Genetics:

A:Gene: yoxD

C:Superfamily: ribitol dehydrogenase; short-chain alcohol dehydrogenase homology

Wed Jun 13 14:51:52 2001

us-09-464-039-7.rpr

Page 7

```

Oy 175 NSMYVLGADEF -GGEIYAVNNLAKETAI-----HAAAMDYGGPIESQCR 219
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 166 MNMSLSLATOYGAHQIRCNNAVAGLILTERLLAKDKCMORHLSRHOQLPRVG-----H 220
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 220 KVDIIADAAYSIFQKPSFGNGV -IDENIL 249
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 PEDVAALVAFELLDSDASFITGGVVCIDGMGL 251
      : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: June 11, 2001, 16:23:48
Job time: 108 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM proteinh - protein search, using sw model

Run on: June 11, 2001, 16:23:30 ; Search time 11.77 Seconds
(without alignments)
1216.550 Million cell updates/sec

Title: US-09-464-039-7

Sequence: 1 MLPTGRLAGCTVFITGASR.....GNMALAIRLEKLMQNMARL 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SWISSPROT_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	256	12.0	238	1	YOXD_BACSU
2	241	11.3	246	1	FABG_THEMA
3	237	11.1	258	1	DHG2_BACSU
4	236	11.0	736	1	DHB4_HUMAN
5	232	10.8	278	1	YALA_RHISN
6	231	10.8	894	1	FOX2_NEUCR
7	224	10.5	285	1	UCPA_ECOLI
8	223	10.4	275	1	YCPI_BRAJA
9	222.5	10.4	271	1	SDRI_PICAB
10	222.5	10.4	320	1	FABG_CUPLA
11	218.5	10.2	246	1	FABG_BACSU
12	218.5	10.2	251	1	Y325_THEMA
13	217	10.1	261	1	DHG_BACSU
14	211.5	9.9	241	1	BUDC_KLETE
15	211	9.9	900	1	FOX2_YEAST
16	210.5	9.8	256	1	BUDC_KLEPN
17	210.5	9.8	735	1	DHB4_RAT
18	210	9.8	258	1	BDHA_RHIME
19	206.5	9.7	285	1	GS33_BACSU
20	206	9.6	592	1	EPHD_MYCTU
21	205.5	9.6	319	1	FABG_ARATH
22	204.5	9.6	253	1	KDUD_ERMCH
23	203	9.5	255	1	HDHA_ECOLI
24	201.5	9.4	261	1	DHG1_BACME
25	201.5	9.4	261	1	DHG2_BACME
26	200.5	9.4	735	1	DHB4_MOUSE
27	200	9.4	257	1	YXJF_BACSU
28	199.5	9.3	247	1	FAGI_SYNY3
29	198.5	9.3	262	1	DHGB_BACME
30	197.5	9.2	906	1	FOX2_CANTR
31	197	9.2	287	1	HETN_ANASP
32	195.5	9.1	253	1	KDUD_ECOLI
33	194	9.1	261	1	DHG2_BACME

ALIGNMENTS

34	193	9.0	250	1	LINX_PSEPA	P50198 pseudomonas
35	193	9.0	270	1	DHMA_FLASI	P22441 flavobacter
36	191.5	9.0	262	1	VER1_ASPEA	P50161 aspergillus
37	189.5	8.9	264	1	STCO_EMENT	O00791 emeritella
38	189	8.8	261	1	DHG4_BACME	P39485 bacillus me
39	189	8.8	261	1	DHG4_BACME	P10528 bacillus me
40	188.5	8.8	262	1	YXBG_BACSU	P46331 bacillus su
41	187	8.7	261	1	DHG3_BACME	P39484 bacillus me
42	181	8.5	236	1	Y019_THEMA	O56318 thermotoga
43	180.5	8.4	258	1	BDHA_ALCEU	O9x6u2 alcaligenes
44	179.5	8.4	275	1	BNZE_PSEPU	P08088 pseudomonas
45	179.5	8.4	275	1	TODD_PSEPU	P13859 pseudomonas

RESULT 1
YOXD_BACSU STANDARD; PRT; 238 AA.
ID YOXD_BACSU
AC P14802;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL OXIDOREDUCTASE IN RTP-PELB INTERGENIC REGION (EC 1.-.-.-)
DE (ORF238).
GN YOXD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=91192601; PubMed=1849493;
RA Ahn K.S., Wake R.G.;
RT "Variations and coding features of the sequence spanning the
RT replication terminus of Bacillus subtilis 168 and W23 chromosomes.";
RL Gene 98:107-112(1991).
RN [2]
RP SEQUENCE OF 62-238 FROM N.A.
RC STRAIN=168;
RX MEDLINE=88040469; PubMed=3118336;
RA Carrigan C.M., Haarsma J.A., Smith M.T., Wake R.G.;
RT "sequence features of the replication terminus of the Bacillus
RT subtilis chromosome."
RL Nucleic Acids Res. 15:8501-8509(1987).
CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDRI FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X06168; CAA29533.1; -
CC EMBL; 299114; CAB13743.1; -
CC PIR; S01270; S01270.
CC HSSP; P19992; 2HSD.
CC Subtilisin; Bg11048; yoxd.
CC Interpro; IPR002196; -
CC DR Pfam; PF00106; adh_short; 1.
CC DR PRINTS; PRO0080; SDRFAMILY.
CC DR PROSITE; PS00061; ADH_SHORT; 1.
CC KW Hypothetical protein: Oxidoreductase.
CC NF_BIND 10 34 NAD OR NADP (BY SIMILARITY).
CC ACT_SITE 155 155 BY SIMILARITY.
CC SEQUENCE 238 AA; 25299 MW; 9CC7ABD1204DF248 CRC64;

```

Query Match          12.0%; Score 256; DB 1; Length 238;
Best Local Similarity 30.1%; Pred. No. 1.6e-11;
Matches 81; Conservative 37; Mismatches 97; Indels 54; Gaps 10;

QY 12 TWFITGASGIGKATLAKAKDGANIVIAKTAQPHKILGTITTAEELEAVGKALC 71
   | ||| ||||| | | | | | | | | | | | | | | | | | | | | | | |
DB 8 TALLITGGGIGRATLALAKKESGNIGLIGRTS-----ANEKVAEEVKALGVAAFA 60

QY 72 IYDVDEQOISAVERKAIRKFGIDILVNNASASISLTNTLPTKRLDMMVNTFGTYL 131
   ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 61 AADVADQDVQNAVQVQRKQDIDILINNAGSKRGFLDLSADWEMNIIQVNLMGVYH 120

QY 132 ASKACIPYLKSKVAHIILNISPPNLNPPVFKQHCAYTIATKGSMTV-VLGAEEF----- 186
   : : | : : | : | | | | | | | | | | | | | | | | | | | |
DB 121 VTRAVLPKEMIERKAGDIINISSTAG-----ORGAATVSAYSASKFAVLGLTESLMQEV 173

QY 187 -GGETAVNMLPKRTAHTHTAMDLGGPGTESQCRKVDIADAAVSTFQPKSFPTGNFVID 245
   | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 174 RHNHTRVSLTPS---TVASDM---STELN-----LTD-----GN---P 203

QY 246 ENLKEEGEENEDVYAIKPGHPLOPDFL 274
   | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 204 EKVMQPEDLAEMYVAQLK-----LDPRIF 228

RESULT 2
FABG_THEMEA STANDARD; PRT; 246 AA.
AC 09X248;
DT 30-MAY-2000 (Rel. 39, Last Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 3-OXOACYL-[ACYL-CARRIER PROTEIN] REDUCTASE (EC 1.1.1.100) (3-KETOACYL-
ACYL CARRIER PROTEIN REDUCTASE).
GN FABG OR TM1724.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxId=23356;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwin M.L., Dodson R.J.,
RA Hafe D.H., Hickey E.K., Peterson J.D., Nelson M.C., Ketchum K.A.,
RA McDonald L., Uteerback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RL genome sequence of Thermotoga maritima.";
CC CC
CC -1- CATALYTIC ACTIVITY: (3R)-3-HYDROXYACYL-[ACYL-CARRIER PROTEIN] +
CC NADP(+) -> 3-OXOACYL-[ACYL-CARRIER PROTEIN] + NADPH.
CC -1- PATHWAY: FIRST REDUCTION STEP IN THE FATTY ACID BIOSYNTHESIS
CC PATHWAY.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SCR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AE001811; AAD36790.1; -
CC TIGR: TM1724; -
CC InterPro: IPR002198; -
CC InterPro: IPR002347; -
CC InterPro: IPR002424; -
CC Pfam: PF00106; adh_short; 1.
CC Pfam: PF00678; adh_short_C2; 1.

```

```

DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDRHD.
DR PRINTS: PR01167; INSADHFAMILY.
KW Fatty acid biosynthesis; Oxidoreductase; NADP.
FT NP_BIND 10 34 NADP (BY SIMILARITY).
FT ACT_SITE 154 154 BY SIMILARITY.
SQ SEQUENCE 246 AA; 26401 MW; 8C08904D8209142 CRC64;

Query Match          11.3%; Score 241; DB 1; Length 246;
Best Local Similarity 29.5%; Pred. No. 1.9e-10;
Matches 69; Conservative 41; Mismatches 76; Indels 48; Gaps 8;

QY 7 RLAGCTVFTGASRGIGKATLAKAKDGANIVIAKTAQPHKILGTITTAEELEAVG 66
   || | ||||| | | | | | | | | | | | | | | | | | | | | |
DB 2 RLEGKVLITGAASGIGKATLTLLFQEGATVIAGDISKE-----NDSLVKEAGLPG 54

QY 67 KALPCIVDVDEQOISAVERKAIRKFGIDILVNNASASISLTNTLPTKRL-----DLA 121
   | : | : | | | | | | | | | | | | | | | | | | | |
DB 55 KVDPPVLTNTDRDQIKVEYKQKGRIDVLVNNAI-----TRDALLVRKKEEDMDAV 109

QY 122 MNVNTFGTYLASKACIPYLKSKVAHIILNISPPNLN-PPVFKQHCAYTIATKGSMTV 179
   ||| | | : | : | : | : | : | : | : | : | : | : | : |
DB 110 INVNLKGVNTQMVVPIKORNGSIVAVSSVGIYGNP-----GOTNAAKAG-----V 161

QY 180 LGMAEEFKGEIA-----VNALMP-----KTAHTAAMDMLGSP 212
   | : | : | | | | | | | | | | | | | | | | | | | |
DB 162 IGMTKWAKELAGRINRVANAVAPGFIETPTETKLEPKARETALSRIPLGRFGKP 215

RESULT 3
DHG2_BACSU STANDARD; PRT; 258 AA.
AC DHG2_BACSU
ID P80869;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLUCOSE 1-DEHYDROGENASE II (EC 1.1.1.47) (GLCDH-II) (GDH-II) (GENERAL
DE STRESS PROTEIN 74) (GSP74).
GN YCDF.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OX NCBI_TaxId=1423;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kumano M., Tamakoshi A., Yamane K.;
RT "A 32 kb nucleotide sequence from the region of the lincomycin-
RT resistance gene (22-25 degree) of the Bacillus subtilis chromosome and
RT identification of the site of the lin-2 mutation.";
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
CC CC
CC SEQUENCE OF 1-13.
CC STRAIN=ISS8;
CC MEDLINE=97443988; PubMed=9298659;
CC Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
CC Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
CC -1- CATALYTIC ACTIVITY: BETA-D-GLUCOSE + NAD(P)(+) -> D-GLUCONO-DELTA-
CC LACTONE + NAD(P)H.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
CC LIMITATION AND OXYGEN LIMITATION.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SCR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```



```

Db 174 LANSLAIEGRKSNHCNTIAPNAGSRMTQTVMEDELVEALKEPYVAPLYLWLCHESCEEN 233
Oy 191 -----AVNALMPK----- 198
Db 234 GGLFEVAGMICKLWERTLGAIVROKNHPMPREAVKANMKKICDEFENASKPOSIDESTG 233
Oy 199 -----PAHTA----- 204
Db 294 SIEVLKIDSEGVASANTSRATSGAGAIQKLPFSYATLELAIMYALGVGA 353
Oy 205 -----AMDILG-----POIESQCRV----- 221
Db 354 SIKDPKDLFTYEGSSDFECLPTFGVIIGQKSMGGSLAEIPGLSINFAVLHGEQYLEL 413
Oy 222 -----DIADA-----AYSIFQK----- 224
Db 414 YKPLPAGKLKCAVAVADVLDKSGSVIIMDYVSYSEKELICHNOPSFLVSGSGGGR 473
Oy 235 -----PKSFTGNFV-----IDENILKE 251
Db 474 TSDKVVAVAAIPNRPDAVLTDTTSLNOALYRLSGDMWPLHIDPWFASLAGFDPKIL-- 531
Oy 252 EGIENE-----DVAIRK-----PGHLPDPDFLEDEPEAVSKVYE 286
Db 532 HGLCTGFESARRVLOQPADNDVSRFAIKARFAPYFGOTLQTEMKKGNRIHPOTKVO 591
Oy 287 STGAVEPEFEKQLQLP-----KPRGAVEEFT-----RIVDSLSDDVYKATQAI 332
Db 592 ETCGD-VISNAVYDLAPFTSGTSAKTPSEGGKLOSTFVEEIGRLKD-IGPEVVKVAVN 649
Oy 333 YLEFLS--GEDGTWFLDLKSKGNGVGEPSQADVNSMTTDDVVKMFSGKLPTMAF 390
Db 650 FEHHTIRKGNIGAKKMTIDIKSGSKVYQGPAGKADTTIILSDEDMEVYGLKLDQKAF 709
Oy 391 MSGKTKINGMALAKLEKLM 411
Db 710 FGSRLLKARGNIMLSOKLQML 730

RESULT 5
Y4LA_RHISN
ID Y4LA_RHISN STANDARD: PRT: 278 AA.
AC P55541:
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 01-NOV-1997 (rel. 35, Last annotation update)
DE PUTATIVE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE Y4LA (EC 1.-.-.-).
GN Y4LA.
OS Rhizobium sp. (strain NGR234).
OC Plasmid sym PNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97305956; PubMed-9163424;
RA Freilberg C.A., Fellay R., Bailroch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000082; AAB91754.1; -
CC HSP: P29132; IDFI.

```

```

DR InterPro: IP002398; -.
DR InterPro: IP002347; -.
DR Pfam: PF00106; adh_shortc.1.
DR Pfam: PF00678; adh_shortc2.1.
DR PRINTS: PR00080; SDRFAMILY.
DR PRINTS: PR00081; GDHRDH.
DR PROSITE: PS00061; ADH_SHORT.1.
KW Hypothetical protein; Oxidoreductase; Plasmid.
FT NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).
FT ACT_SITE 159 159 BY SIMILARITY.
SQ SEQUENCE 278 AA; 28743 MW; 1D0105625DE9DF2B CRC64;

Query Match 10.8%; Score 232; DB 1; Length 278;
Best Local Similarity 29.5%; Pred. No. 16-09;
Matches 80; Conservative 34; Mismatches 107; Indels 50; Gaps 8;

Oy 7 RLACGTVFTTGASRGIGKALKAANDGANIYA-----AKTQPPKLLGTITTA 58
Db 3 RFEKVAVYTGAGAGIGKACALAIAREGRRVAVADGDGSAIACDHO----- 49
Oy 59 EETEAAGKALPCIVDVRDEQOISAVERAIKKGSGIDILVNNASISLT----NTLTP 114
Db 50 --IAAEGNALAMADIADQAVAALEFTRAERHFGVDLLVNNASMHILTPDRALDLD 107
Oy 115 TKRLDLMNVNTRGYLASKACIPYLKSKVAHILNISPLNLPVWRQHCAYTIATG 174
Db 108 LAWDDQTMATNLGTLCLCRQALPRMIARGGAIYVNSSCGQIS--GDPAQTSYAVSKAA 165
Oy 175 MSNYYGVMAEEF-KGEIYNALMPKTAI-----HTAAMDILGSGCIESQGR 219
Db 166 MNNLSASLATQGHADIRCAVAPGLMTERLAKLDCKMKRHLRHQLPRVG-----H 220
Oy 220 KVDIIADAAVSIPOKPKSFTGNFV-IDENIL 249
Db 221 PEDVAAALVAFILSDDSFTTGCVYCIDGGWL 251

RESULT 6
FOX2_NEUCR
ID FOX2_NEUCR STANDARD: PRT: 894 AA.
AC 001373:
DT 15-DEC-1998 (rel. 37, Created)
DT 15-DEC-1998 (rel. 37, Last sequence update)
DT 30-MAY-2000 (rel. 39, Last annotation update)
DE PEROXISOMAL HYDRAEASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL
DE BETA-OXIDATION PROTEIN) (MFP) (INCLUDES: 2-ENOYL-COA HYDRAEASE
DE (EC 4.2.1.-); D-3-HYDROXYACYL COA DEHYDROGENASE (EC 1.1.1.-)).
GN FOX-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 497-525 AND 604-619.
RX STRAIN-74-OR8-1A / DSM 1258;
RA MEDLINE-95231521; PubMed-7715608;
RA Fossa A., Beyer A., Pfister E., Wenzel B., Kunau W.-H.;
RT "Molecular cloning, sequencing and sequence analysis of the fox-2 gene
RT of Neurospora crassa encoding the multifunctional beta-oxidation
RT protein.";
RL Mol. Genet. 247:95-104(1995).
CC -1- FUNCTION: SECOND TRIFUNCTIONAL ENZYME ACTING ON THE BETA-OXIDATION
CC PATHWAY FOR FATTY ACIDS, POSSESSING HYDRAEASE-DEHYDROGENASE-
CC EPIMERASE ACTIVITIES. CONVERTS TRANS-2-ENOYL-COA VIA D-3-
CC HYDROXYACYL-COA TO 3-KETOACYL-COA.
CC -1- PATHWAY: BETA-OXIDATION PATHWAY.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CATALASE-FREE MICROBODIES.
CC -1- DOMAIN: CONTAINS TWO SDR DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X80052: CA56355.1; -
 DR InterPro: IPR002198; -
 DR InterPro: IPR002347; -
 DR InterPro: IPR002539; -
 DR Pfam: PF01575: Maoc_dehydratase_1.
 DR Pfam: PF00106: adh_short; 2.
 DR PRINTS: PR00080: SDRFAMLY.
 DR PRINTS: PR00081: GDRHD.
 DR PROSITE: PS00061: ADH_SHORT; 1.
 DR Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
 DR Lyase; Isomerase; Repeat.
 FT DOMAIN 6 230 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT NP_BIND 13 523 NAD (BY SIMILARITY).
 FT ACT_SITE 164 164 BY SIMILARITY.
 FT NP_BIND 318 342 NAD (BY SIMILARITY).
 FT ACT_SITE 458 458 BY SIMILARITY.
 SQ SEQUENCE 894 AA: 96326 MW: 29758220D397AA88 CRC64:

Query Match 10.8%; Score 231; DB 1; Length 894;
 Best Local Similarity 24.2%; Pred. No. 5,6e-09;
 Matches 114; Conservative 64; Mismatches 167; Indels 126; Gaps 20;

QY 7 RIAGCTVITGSKRIGRAIAKAKDGANIY-----AAKTAQHPKILGTI 54
 DB 6 REDGVVVVTVGGGIGAYCYFEGSRGASVYVNDIGASFKEGNSKTKAD----- 56
 QY 55 YTAAGEIENAVGKALPCIVDVRDEQOISAAREKAKKFGGIDILVNNNSA-----ISLTNTL 111
 DB 57 -VYVNEIENAGKAVANNDVSENGDKI---ITAIKEFERIDILINNAGILMDISEKKNK 112
 QY 112 DTPTRRLDMNVNTRGTYLAKACIPYLKSKVAHILNISPLNL--NPVFKQHCAYT 169
 DB 113 D---EDMDLIFKVVHKYSKTARAAAPYFRKQKFGRVINTASAGLFGN---FGQ-ANYS 165
 QY 170 IAKYMSKTVLGMAGEE-EKGEIYAVNALMPKTAIHTAADMELGCGPGIESCKRVYDIADRA 228
 DB 166 AAKIGMVGFTETLAKEGKYNIISNVIAPIA--SRMTETVMPDILLAKMEVWVPLVA 223
 QY 229 YSIFQPKPSFTGNFY-----IDEN-----IKKEGIEHNPDYVA 261
 DB 224 VLVIHKNNTSETGISIEVGGGHHVAKLRWERSGLLKADSEYTPGAILIKMDVDYDPS--- 280
 QY 262 IKRPHLOPDEFELDEYPEAVSKKVESTGAVPEFKEE-----KIQLOPPRSGA 309
 DB 281 -NPQPTGPNDFLALLESIRKGPNDPEKRVDFKGRVALVTGGAGIGAYCLARARAGA 339
 QY 310 VEETRIYKDSIS--DDVYKATQATITLFEISGEDGTWFLDLKSKGN--VGIEPSDQADV 367
 DB 340 SV-----VYNDLVNPDVYV-----EIKKMGKAVGAFKFSLEDGDA 375
 QY 368 VMSMTIDPFVKFKSKLKPMAFMGSKLIRKGNMALIRKLEKLMQNMARL 418
 DB 376 VKKAIDAF-----GRVDIVYVNA--GILIRKAFINMDSL 409

RESULT 7
 ID UCPI_ECOLI STANDARD: PRT: 285 AA.
 AC P37440: P77442: P76863: P77140:
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE OXIDOREDUCTASE UCPIA (EC 1.-.-.-).

GN UCPIA.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97349980; PubMed=9205837;
 RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
 RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mizunashi N.,
 RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
 RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubaram S.,
 RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
 RA Yamagata S., Horiuchi T.;
 RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
 RT K-12 genome corresponding to 50,0-68.8 min on the linkage map and
 RT analysis of its sequence features."
 RL DNA Res. 4:91-113(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=97385354; PubMed=9241368;
 RA Sirko A., Weglenska A., Hryniewicz M.M., Hulanicka D.M.;
 RT "Characterization of the Escherichia coli gene encoding a new member
 RT of the short-chain dehydrogenase/reductase (SDR) family."
 RL Acta Biochim. Pol. 44:153-157(1997).
 RN [4]
 RP SEQUENCE OF 202-285 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=90264335; PubMed=2188959;
 RA Hryniewicz M.M., Sirko A., Palucha A., Boeck A., Hulanicka D.M.;
 RT "Sulfate and thiosulfate transport in Escherichia coli K-12:
 RT identification of a gene encoding a novel protein involved in
 RT thiosulfate binding."
 RL J. Bacteriol. 172:3358-3366(1990).
 RN [5]
 RP IDENTIFICATION.
 RX MEDLINE=95075659; PubMed=7984428;
 RA Borodovsky M., Rudd K.E., Koonin E.V.;
 RT "Intrinsic and extrinsic approaches for detecting genes in a
 RT bacterial genome."
 RL Nucleic Acids Res. 22:4756-4767(1994).
 CC -I- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE000330: AAC75479.1; -
 DR EMBL: D90872: BAA16309.1; -
 DR EMBL: D90871: BAA16300.1; -
 DR EMBL: X99908: CA681801.1: ALT_INT.
 DR EMBL: M32101: -; NOT_ANNOTATED_CDS.
 DR HSSP: P14061: IFDV.
 DR Ecogene: EG12133: ucpiA.
 DR InterPro: IPR002198; -
 DR InterPro: IPR002347; -

DR Pfam: PF00106; adh_short; 1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRDH.
 DR PROSITE: PS00061; ADH_SHORT; FALSE_NEG.
 KM Oxidoreductase.
 FT CONFLICT 175 175 T -> Q (1N REF. 3).
 SQ SEQUENCE 285 AA; 30581 MW; 8B861B6F3BC897A5 CRC64;

Query Match 10.5%; Score 224; DB 1; Length 285;
 Best Local Similarity 30.1%; Pred. No. 3.8e-09;
 Matches 59; Conservative 42; Mismatches 79; Indels 16; Gaps 4;

QY 6 GRLAGCTVFITGASRGIGAKALAKDGANIYAAQPHPKLLGTITYAAEIEAVG 65
 DB 24 GKLTGTALTALTGLOGEGSIARTFARHGANLILDISPE-----LEKLDELQGRG 75
 QY 66 GKALPCIVDYREDOQISAVERAKKFGGIDILVNNASAIISLTNTLDTPTKRLDLMNNV 125
 DB 76 HRCYAVADVDRPASVAAAKRAKEGRIDILVNNAGVCLGSLFDMSDDDDFHIDIN 135
 QY 126 TPTYLAKSKACIPLYL--KRSVAHILNISPLNLPVWFKCHCAVTIAKYGSMVYLGW 182
 DB 136 IKGVNVTYRAVLEPMIARKDRIYMMSVYGDVADP-----GETATATLTAALVGLTKSL 191
 QY 183 AEEF-KGEIYVNAALMP 197
 DB 192 AVEYAGSGIRVNAICP 207

RESULT 8
 YCPL_BRAJA STANDARD; PRT; 275 AA.
 ID YCPL_BRAJA

AC 045219;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE PROBABLE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE (EC 1.-.-.-).
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Bradyrhizobium group; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RA Tully R.E., Keister D.L.;
 RT "Cloning and mutagenesis of a cytochrome P-450 locus from Bradyrhizobium japonicum that is expressed anaerobically and symbiotically."
 RL Appl. Environ. Microbiol. 59:4136-4142(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=98322110; PubMed=9655913;
 RA Tully R.E., van Berkum P., Lovins K.W., Keister D.L.;
 RT Identification and sequencing of a cytochrome P450 gene cluster from Bradyrhizobium japonicum."
 RL Biochim. Biophys. Acta 1398:243-255(1998).
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: U12678; AAC28892.1; -
 DR HSSP: P08074; ICYD.
 DR InterPro: IPR002198; -

DR Pfam: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 KM Hypothetical protein; Oxidoreductase.
 FT NP_BIND 10 34 NAD OR NADP (BY SIMILARITY).
 FT ACT_SITE 159 159 BY SIMILARITY.
 SQ SEQUENCE 275 AA; 28827 MW; 312E7070C404DF86 CRC64;

Query Match 10.4%; Score 223; DB 1; Length 275;
 Best Local Similarity 33.2%; Pred. No. 4.3e-09;
 Matches 68; Conservative 27; Mismatches 80; Indels 30; Gaps 5;

QY 6 GRLAGCTVFITGASRGIGAKALAKDGANIYA-----AKTAQPHPKLLGTITYA 57
 DB 2 GREGKVAAYVTGAGAGIGKACALAIAREGGVAVVDIGSAAIACIAQA----- 49
 QY 58 AEEIEAVGKALPCIVDYREDOQISAVERAKKFGGIDILVNNASAIISLT---NTLDT 113
 DB 50 ---TAAEGHALLALADIDADQAVALFETAEHFGGVLDLVNNASAMHLLTPDRAILEL 106
 QY 114 PTKRLDLMNNVNTRGTYLASKACIPLYLKKSVAHILNISPLNLPVWFKCHCAVTIAKY 173
 DB 107 ELAVNQOTMARNLKGLTLCROAIPIRIARAGGAIYVNMSSCGLS--GDTALTYSASKA 164
 QY 174 GMSYVVGMAEEF-KGEIYVNAALMP 197
 DB 165 AMNMSSSLATQYGHQAIQCNAAVAP 189

RESULT 9
 SDR1_PICAB STANDARD; PRT; 271 AA.
 ID SDR1_PICAB

AC 008632;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE SHORT-CHAIN TYPE DEHYDROGENASE/REDUCTASE (EC 1.-.-.-).
 OS Picea abies (Norway spruce) (Picea excelsa).
 OC Coniferopsida; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Picea.
 OX NCBI_TaxID=3329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=94120027; PubMed=8290650;
 RA Bauer S., Galliano H., Pfeiffer F., Messner B., Sandermann H., Ernst D.;
 RT "Isolation and characterization of a cDNA clone encoding a novel short-chain alcohol dehydrogenase from Norway spruce (Picea abies L. karst)."
 RL Plant Physiol. 103:1479-1480(1993).
 CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES (SDR) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: X74115; CAA52213.1; -
 DR HSSP: Q12634; IYBV.
 DR InterPro: IPR002198; -
 DR InterPro: IPR002347; -
 DR Pfam: PF00106; adh_short; 1.
 DR Pfam: PF00678; adh_short_C2; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRDH.
 DR PROSITE: PS00061; ADH_SHORT; 1.
 KM Oxidoreductase.
 FT NP_BIND 25 49 NAD OR NADP (BY SIMILARITY).


```

OC Bacillus/Staphylococcus group; Bacillus.
OX NCB1_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=86168021; PubMed=3082854;
RT Lampe K.A., Urrant B., Chaudhry G.R., Rameley R.F., Rudikoff S.;
RT "Characterization of the developmentally regulated Bacillus subtilis
RT glucose dehydrogenase gene."
RL J. Bacteriol. 166:238-243(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124189; PubMed=8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis chromosome:
RT determination of the sequence of a 146 kb segment and identification
RT of 113 genes."
RL Microbiology 142:3047-3056(1996).
CC -1 CATALYTIC ACTIVITY: BETA-D-GLUCOSE + NAD(P)(+) = D-GLUCONO-DELTA-
CC LACTONE + NAD(P)H.
CC -1 SUBUNIT: HOMOTETRAMER.
CC -1 INDUCTION: IT IS INDUCED AT STAGE III OF THE SPOULATION.
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M12276; AAA22463.1;
CC EMBL: D50453; BAA09024.1;
CC EMBL: 299106; CAB12201.1;
CC PIR: S36090; S36090.
CC HSSP: P19992; 2HSD.
CC Sublist: BG10545; gdh.
CC InterPro: IPR002198;
CC InterPro: IPR002347;
CC Pfam: PF00106; adh_short; 1.
CC Pfam: PF00678; adh_short_C2; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PRINTS: PR00081; GDHHDH.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Oxidoreductase; NADP; Sporulation.
CC NP_BIND 11 35 NADP (BY SIMILARITY).
CC ACT_SITE 158 158 BY SIMILARITY.
CC CONFLICT 148 150 EVI -> AE (IN REF. 1).
CC SEQUENCE 261 AA; 28090 MW; 5894C17DB8F14965 CRC64;
SQ

```

```

Query Match 10.1%; Score 217; DB 1; Length 261;
Best Local Similarity 31.6%; Pred. NO. 1.le-08;
Matches 65; Conservative 42; Mismatches 83; Indels 16; Gaps 7;

```

```

OY 1 MNPNTGRLACGVFTTGASRGIGKAIKAAKGDANIVIAAKTADPHKLLGTYTAEE 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MYPD--LKGKVAITGAASGLGKAMAIRFGKQAKVIVNYNSKQDPN-----EVKEE 51
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 61 IEAVGKALPCIVDVRDEQOISAAVEKAIRKFGIDILVNNASALSJTLDTPTKRLD 120
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 52 VIKAGEAAVVOGDVTKEDVKNIVOTAIKFGTLDIMNNGLEPNVPSHEMPLKDWK 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 121 MNMNVTRGYTLASKACIPILKSKV-AHLINTSPPLNLPVFKOHCATYIAKYGSMYV 179
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 112 VIGTNTGAFLSRBAIKFVENDIKGNVYNSVHEVLPWLPVH-YAAKSGGIKIMT 169
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 180 LGMAEF--KGEIIVNALMPKTAIHT 203
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 170 ETIALEYAPKG-IRVNNIGP-GAINT 193
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 14
ID BUDC_KLETE STANDARD: PRT: 241 AA.
AC 004520;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ACETOIN(DIACETYL) REDUCTASE (EC 1.1.1.5) (ACETOIN DEHYDROGENASE) (AR).
GN BUDC.
OS Klebsiella terrigena.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Klebsiella.
OX NCB1_TaxID=577;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VTT-E-74023;
RX MEDLINE=93186707; PubMed=8444801;
RA Blomqvist K., Nikkola M., Lehtovaara P., Suikko M.-L.,
RA Airaksinen U., Straby K.B., Knowles J.K.C., Penttilae M.E.;
RT "Characterization of the genes of the 2,3-butanediol operons from
RT Klebsiella terrigena and Enterobacter aerogenes."
RL J. Bacteriol. 175:1392-1404(1993).
CC -1 CATALYTIC ACTIVITY: ACETOIN + NAD(+) = DIACETYL + NADH.
CC -1 SUBUNIT: HOMOTETRAMER.
CC -1 SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
CC (SDR) FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: L04507; AAA25056.1;
CC PIR: EA7068; E47069.
CC HSSP: P19992; 2HSD.
CC InterPro: IPR002198;
CC Pfam: PF00106; adh_short; 1.
CC PRINTS: PR00080; SDRFAMILY.
CC PROSITE: PS00061; ADH_SHORT; 1.
CC Oxidoreductase; NAD.
CC NP_BIND 6 30 NAD (BY SIMILARITY).
CC ACT_SITE 152 152 BY SIMILARITY.
CC SEQUENCE 241 AA; 25313 MW; D85801C56931BF04 CRC64;
SQ

```

```

Query Match 9.9%; Score 211.5; DB 1; Length 241;
Best Local Similarity 30.1%; Pred. NO. 2.4e-08;
Matches 68; Conservative 34; Mismatches 71; Indels 53; Gaps 10;

```

```

OY 15 ITGASRGIGKATLAKKAAKGDANIVIAAKTADPHKLLGTYTAEEIEAVGKALPCIVD 74
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 7 VIGAGGIGKATLAKLRVYKGFVALADYND-----AATVAAEINAGGRAVAIKYD 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 75 VDEQOISAAVEKAIRKFGIDILVNNASALSJTLDTPTKRL-DLMNNTVTRGYTLAS 133
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 60 VSRDQVFAAVDAQARKALGFPVIVNNA-GIAPSPFIESITBEIVDRVYNINWKGIVMCM 118
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 134 KACIPILK-----SKVAHLNTSPPLNLPVFKOHCATYIAKYGSMYVLCM 182
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 119 QAAVEAFKKEGHGKIVNACSOAGHGN--PEL-----AYSSSKFA-----VRGL 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 183 AEEFKEIA-----VNALMP--KTAIHTAAMDMLGPGIESQCKR 220
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 163 TOTARDLAPLGITYVNGPFGIVKTPMAE-----IDQCKR 199
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 15
FOX2_YEAST
ID FOX2_YEAST STANDARD: PRT: 900 AA.

```

AC 002207;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PEROXISOMAL HYDRATASE-DEHYDROGENASE-EPIMERASE (HDE) (MULTIFUNCTIONAL
 BETA-OXIDATION PROTEIN) (MFP) [INCLUDES: 2-ENOYL-COA HYDRATASE
 (EC 4.2.1.-); D-3-HYDROXYACYL COA DEHYDROGENASE (EC 1.1.1.-)].
 GN FOX2 OR YKR009C OR YK108.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
 ON NCBI_taxid=4932;
 RX MEDLINE=93070612; PubMed=1441752;
 RA Desferioheft A., Philippson P.,
 RT "DNA sequencing and analysis of a 24.7 kb segment encompassing
 RT centromere CEN1 of Saccharomyces cerevisiae reveals nine previously
 RT unknown open reading frames.";
 RL Yeast 8:749-759(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92202210; PubMed=1551874;
 RA Hiltunen J.K., Wenzel B., Beyer A., Erdmann R., Fossa A., Kunau W.H.,
 RT "Peroxisomal multifunctional beta-oxidation protein of Saccharomyces
 RT cerevisiae. Molecular analysis of the fox2 gene and gene product.";
 RL J. Biol. Chem. 267:6646-6653(1992).
 CC -1- FUNCTION: SECOND TRIFUNCTIONAL ENZYME ACTING ON THE BETA-OXIDATION
 CC PATHWAY FOR FATTY ACIDS, POSSESSING HYDRATASE-DEHYDROGENASE-
 CC EPIMERASE ACTIVITIES. CONVERTS TRANS-2-ENOYL-COA VIA D-3-
 CC HYDROXYACYL-COA TO 3-KETOACYL-COA.
 CC -1- PATHWAY: BETA-OXIDATION PATHWAY.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: PEROXISOMAL.
 CC -1- DOMAIN: CONTAINS TWO SDR DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
 CC (SDR) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M86456; AAA34779.1; -;
 DR EMBL: X65124; CAA46243.1; -;
 DR EMBL: Z28234; CAA82079.1; -;
 DR PIR: S25322; S25322.
 DR HSSP: P25529; LAHH.
 DR SGD: S0001717; FOX2.
 DR InterPro: IPR002198; -;
 DR InterPro: IPR002539; -;
 DR Pfam: PF01575; Maoc_dehydratase; 1.
 DR Pfam: PF00106; adh_short; 2.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PROSITE: PS00061; ADH_SHORT; 2.
 DR PROSITE: PS00342; MICRORODIES_CTER; 1.
 KW Fatty acid metabolism; Multifunctional enzyme; Oxidoreductase; NAD;
 KW Lyase; Isomerase; Peroxisome; Repeat.
 FT DOMAIN 6 230 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT SITE 319 535 SHORT-CHAIN DEHYDROGENASE LIKE.
 FT NP_BIND 898 900 MICROBODY TARGETING SIGNAL (POTENTIAL).
 FT NP_BIND 13 37 NAD (BY SIMILARITY).
 FT ACT_SITE 165 165 NAD (BY SIMILARITY).
 FT NP_BIND 326 350 NAD (BY SIMILARITY).
 FT ACT_SITE 469 469 BY SIMILARITY.
 SO SEQUENCE 900 AA; 98703 MW; 66FFDDA9C673788 CRC64;

Query Match

9.9% Score 211; DB 1; Length 900;

Best Local Similarity 26.0%: Pred. No. 1,5e-07;
 Matches 109; Conservative 50; Mismatches 147; Indels 114; Gaps 21;
 QY 13 VFETGASRGICATATLAAKDGANIVIAATAQPHPLTGTI-----YTAEEIE 62
 DB 12 VVITGAGGGLCKVYALVYASGAKVYV-----NLTGTLGGSGNSKAADLYVEIK 63
 QY 63 AVGKALPCIVDVDEQOISAAYEKAIRKFGIDILVNNASAIJLNTLDTPTKRLDLM 122
 DB 64 KAGGIAVANSDVNGE--KIETAIKEFGYVYLLNNGILRDVSFAKWTREFAVY 121
 QY 123 NVNTRGTYLAKACIPYLKSKVAHIIISPLNL--NPWFKQHCAYTAKYMSYVL 180
 DB 122 DVHLTGKYLKSRAMPYMRSGQFGRITINTASPAFLFEN---FGQ-ANYSAKMGL---V 173
 QY 181 GMAEEF-----KGEIYVNALMPKTAIHTAAMDMLGGPISQCKVLIIDAAV----- 229
 DB 174 GLAETLAKKGAKNINVNSIAP--LARSMTENVLPPIILKQLGPEKIVPLVYLTHEST 231
 QY 230 ----SIFQPKSFTGN-----FVIDENILKEGIEN---FDVYAIKPGHPLOPD 271
 DB 232 KVSNSIFELAGFEGQLRMERSGQIFNPDKTYTPPAILNKKWEITDYDRKPNKTOHP 291
 QY 272 FFLDEYEPVASKVYESGAVPEFEKEKLOLPKRSQAVETFRIVKDSLDVYKATQA 331
 DB 292 YQLSDYNDLIT-----NAKRL---PPNEGQSVK---IK-SLCNKVYVYV-- 328
 QY 332 IYLFELSGEDG-----TWFL-----DLKSKG-----NWGYGEPG---DQADV 368
 DB 329 -----GAGGGLCKSHAIWPARYGAKVYVNDIDPFSYVEIKKLKGEGALIPDSHDV 381

Search completed: June 11, 2001, 16:25:47
 Job time: 137 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2001, 16:20:50 ; Search time 22.06 Seconds

(without alignments)
1063.147 Million cell updates/sec

Title: US-09-464-039-7

Perfect score: 2139

Sequence: 1 MLENTGRLAGCTVFITGASR.....GNMALIKLEKLNQNMARL 418

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

A.GeneSeq_0401.*
1: /SID56/gcgdata/geneSeq/geneSeq/AA1980.DAT.*
2: /SID56/gcgdata/geneSeq/geneSeq/AA1981.DAT.*
3: /SID56/gcgdata/geneSeq/geneSeq/AA1982.DAT.*
4: /SID56/gcgdata/geneSeq/geneSeq/AA1983.DAT.*
5: /SID56/gcgdata/geneSeq/geneSeq/AA1984.DAT.*
6: /SID56/gcgdata/geneSeq/geneSeq/AA1985.DAT.*
7: /SID56/gcgdata/geneSeq/geneSeq/AA1986.DAT.*
8: /SID56/gcgdata/geneSeq/geneSeq/AA1987.DAT.*
9: /SID56/gcgdata/geneSeq/geneSeq/AA1988.DAT.*
10: /SID56/gcgdata/geneSeq/geneSeq/AA1989.DAT.*
11: /SID56/gcgdata/geneSeq/geneSeq/AA1990.DAT.*
12: /SID56/gcgdata/geneSeq/geneSeq/AA1991.DAT.*
13: /SID56/gcgdata/geneSeq/geneSeq/AA1992.DAT.*
14: /SID56/gcgdata/geneSeq/geneSeq/AA1993.DAT.*
15: /SID56/gcgdata/geneSeq/geneSeq/AA1994.DAT.*
16: /SID56/gcgdata/geneSeq/geneSeq/AA1995.DAT.*
17: /SID56/gcgdata/geneSeq/geneSeq/AA1996.DAT.*
18: /SID56/gcgdata/geneSeq/geneSeq/AA1997.DAT.*
19: /SID56/gcgdata/geneSeq/geneSeq/AA1998.DAT.*
20: /SID56/gcgdata/geneSeq/geneSeq/AA1999.DAT.*
21: /SID56/gcgdata/geneSeq/geneSeq/AA2000.DAT.*
22: /SID56/gcgdata/geneSeq/geneSeq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244	11.4	258	21	Y56815
2	232.5	10.9	18	18	W16329
3	224	10.5	240	16	R61477
4	224	10.5	247	16	R77866
5	218.5	10.2	248	21	Y54422
6	217	10.1	261	21	Y96271
7	217	10.1	261	21	Y54424
8	206.5	9.9	900	20	Y30338
9	206.5	9.7	315	17	R89323
10	206.5	9.7	315	17	R89322
11	203.5	9.5	313	21	B42640

Result No.	Score	Query Match	Length	ID	Description
12	202.5	9.5	313	20	Y27004
13	199.5	9.3	261	13	R27757
14	198.5	9.3	261	11	R03846
15	198	9.3	251	21	Y55986
16	197.5	9.2	903	20	Y28713
17	197.5	9.2	906	20	Y28702
18	197.5	9.2	906	20	Y28712
19	197	9.2	255	13	R22993
20	196.5	9.2	261	11	R04044
21	196.5	9.2	261	13	R24018
22	193.5	9.0	285	21	G40333
23	193.5	9.0	307	21	G40332
24	191.5	9.0	246	21	B15707
25	191.5	9.0	262	20	Y08329
26	191.5	9.0	298	21	G50312
27	191.5	9.0	315	21	G50311
28	190.5	8.9	298	21	G24787
29	190.5	8.9	316	21	G24786
30	190	8.9	285	21	G12862
31	190	8.9	307	21	G12861
32	189.5	8.9	186	19	W38474
33	189.5	8.9	243	19	W80670
34	189.5	8.9	243	21	B15706
35	189	8.8	272	21	B10740
36	189	8.8	340	21	B10741
37	188.5	8.8	292	21	Y98000
38	186.5	8.7	260	17	R94617
39	186	8.7	261	9	P80590
40	186	8.7	261	13	R27756
41	185.5	8.7	277	21	B29473
42	185	8.6	261	9	P80063
43	183	8.5	276	16	R66734
44	181.5	8.5	285	21	G12202
45	181.5	8.5	285	21	G47575

ALIGNMENTS

Result No.	Score	Query Match	Length	ID	Description
1	244	11.4	258	21	Y56815
2	232.5	10.9	18	18	W16329
3	224	10.5	240	16	R61477
4	224	10.5	247	16	R77866
5	218.5	10.2	248	21	Y54422
6	217	10.1	261	21	Y96271
7	217	10.1	261	21	Y54424
8	206.5	9.9	900	20	Y30338
9	206.5	9.7	315	17	R89323
10	206.5	9.7	315	17	R89322
11	203.5	9.5	313	21	B42640

Result No.	Score	Query Match	Length	ID	Description
12	202.5	9.5	313	20	Y27004
13	199.5	9.3	261	13	R27757
14	198.5	9.3	261	11	R03846
15	198	9.3	251	21	Y55986
16	197.5	9.2	903	20	Y28713
17	197.5	9.2	906	20	Y28702
18	197.5	9.2	906	20	Y28712
19	197	9.2	255	13	R22993
20	196.5	9.2	261	11	R04044
21	196.5	9.2	261	13	R24018
22	193.5	9.0	285	21	G40333
23	193.5	9.0	307	21	G40332
24	191.5	9.0	246	21	B15707
25	191.5	9.0	262	20	Y08329
26	191.5	9.0	298	21	G50312
27	191.5	9.0	315	21	G50311
28	190.5	8.9	298	21	G24787
29	190.5	8.9	316	21	G24786
30	190	8.9	285	21	G12862
31	190	8.9	307	21	G12861
32	189.5	8.9	186	19	W38474
33	189.5	8.9	243	19	W80670
34	189.5	8.9	243	21	B15706
35	189	8.8	272	21	B10740
36	189	8.8	340	21	B10741
37	188.5	8.8	292	21	Y98000
38	186.5	8.7	260	17	R94617
39	186	8.7	261	9	P80590
40	186	8.7	261	13	R27756
41	185.5	8.7	277	21	B29473
42	185	8.6	261	9	P80063
43	183	8.5	276	16	R66734
44	181.5	8.5	285	21	G12202
45	181.5	8.5	285	21	G47575

Result No.	Score	Query Match	Length	ID	Description
12	202.5	9.5	313	20	Y27004
13	199.5	9.3	261	13	R27757
14	198.5	9.3	261	11	R03846
15	198	9.3	251	21	Y55986
16	197.5	9.2	903	20	Y28713
17	197.5	9.2	906	20	Y28702
18	197.5	9.2	906	20	Y28712
19	197	9.2	255	13	R22993
20	196.5	9.2	261	11	R04044
21	196.5	9.2	261	13	R24018
22	193.5	9.0	285	21	G40333
23	193.5	9.0	307	21	G40332
24	191.5	9.0	246	21	B15707
25	191.5	9.0	262	20	Y08329
26	191.5	9.0	298	21	G50312
27	191.5	9.0	315	21	G50311
28	190.5	8.9	298	21	G24787
29	190.5	8.9	316	21	G24786
30	190	8.9	285	21	G12862
31	190	8.9	307	21	G12861
32	189.5	8.9	186	19	W38474
33	189.5	8.9	243	19	W80670
34	189.5	8.9	243	21	B15706
35	189	8.8	272	21	B10740
36	189	8.8	340	21	B10741
37	188.5	8.8	292	21	Y98000
38	186.5	8.7	260	17	R94617
39	186	8.7	261	9	P80590
40	186	8.7	261	13	R27756
41	185.5	8.7	277	21	B29473
42	185	8.6	261	9	P80063
43	183	8.5	276	16	R66734
44	181.5	8.5	285	21	G12202
45	181.5	8.5	285	21	G47575

Result No.	Score	Query Match	Length	ID	Description
12	202.5	9.5	313	20	Y27004
13	199.5	9.3	261	13	R27757
14	198.5	9.3	261	11	R03846
15	198	9.3	251	21	Y55986
16	197.5	9.2	903	20	Y28713
17	197.5	9.2	906	20	Y28702
18	197.5	9.2	906	20	Y28712
19	197	9.2	255	13	R22993
20	196.5	9.2	261	11	R04044
21	196.5	9.2	261	13	R24018
22	193.5	9.0	285	21	G40333
23	193.5	9.0	307	21	G40332
24	191.5	9.0	246	21	B15707
25	191.5	9.0	262	20	Y08329
26	191.5	9.0	298	21	G50312
27	191.5	9.0	315	21	G50311
28	190.5	8.9	298	21	G24787
29	190.5	8.9	316	21	G24786
30	190	8.9	285	21	G12862
31	190	8.9	307	21	G12861
32	189.5	8.9	186	19	W38474
33	189.5	8.9	243	19	W80670
34	189.5	8.9	243	21	B15706
35	189	8.8	272	21	B10740
36	189	8.8	340	21	B10741
37	188.5	8.8	292	21	Y98000
38	186.5	8.7	260	17	R94617
39	186	8.7	261	9	P80590
40	186	8.7	261	13	R27756
41	185.5	8.7	277	21	B29473
42	185	8.6	261	9	P80063
43	183	8.5	276	16	R66734
44	181.5	8.5	285	21	G12202
45	181.5	8.5	285	21	G47575

Result No.	Score	Query Match	Length	ID	Description
12	202.5	9.5	313	20	Y27004
13	199.5	9.3			


```
DE Clavulanic acid dehydrogenase sequence.
XX
KM Clavulanic acid; antibiotic; Augmentin.
OS Streptomyces clavuligerus ATCC 27064.
PN WO9503416-A.
XX
PD 02-FEB-1995.
PF 15-JUL-1994; 94MO-EP02346.
PR 24-JUL-1993; 93GB-0015393.
XX
PA (SMK ) SMITHKLINE BEECHAM PLC.
XX
PI Arnell J, Elson SW, Nicholson NH, Woronicki SR;
XX WPI: 1995-075242/10.
DR N-PSTDB; Q67531.
XX
PT New clavulanic acid dehydrogenase from Streptomyces, and related
PI DNA and vectors - used to produce beta-lactamase inhibiting
PT clavulanic acid from new 3-oxoethylidene analogues
XX
PS Claim 2; Fig 1; 36pp; English.
XX
CC A new enzyme is disclosed which has clavulanic acid dehydrogenase
CC activity, the enzyme having an apparent mol. wt. of 28 kd (by SDS
CC PAGE) and incorporating the N-terminal sequence PSLMGSKVALITGASSGIEF.
CC The enzyme is derived from the mycelium of a Streptomyces species,
CC especially S. clavuligerus (e.g. ATCC 27064), S. jumogensis (e.g. ATCC
CC 29864) or S. katusrahmanus (e.g. T272). The present sequence
CC represents the enzyme from S. clavuligerus ATCC 27064.
CC The enzyme is used to catalyse the biosynthesis of clavulanic acid
CC from a precursor clavulanic acid aldehyde. The obtained clavulanic acid
CC is in turn a key ingredient in the antibiotic Augmentin.
SQ Sequence 240 AA:

Query Match 10.5%; Score 224; DB 16; Length 240;
Best Local Similarity 29.5%; Pred. NO. 1.3e-13;
Matches 71; Conservative 40; Mismatches 96; Indels 34; Gaps 8.

Db 8 LAGCTVFETIGARGRGKATALKAADGANIVTAAKTAPHPKLGTGYAAEIEAVGSK 67
| | | | | | | | | | : | | : | | : | | |
5 Iqgkvailltgassigyeataaraaegaavaalaarive-----klrlgdeilaagk 57
: | | | | | | | | | | : | | : | | : | | |

Db 68 ALPCITVDVDEEQOISAAVEKAIRKFGIDIIYNNASAI SLFTNLTDPTRRLDLMNVNR 127
: | | | | | : | | : | | : | | : | | : | |
58 vhlvldvdadrgvdaavaastvealgydlilvnngimllpgvedatdtcwrmidtnl 117
: | | | | | : | | : | | : | | : | | : | |

Db 128 GTYLASKACIPLPKLSK--VAHILNISPPMLNPWFVKQCHCAITYTAKYMSKYVLGMABE 185
: | : | : | : | | : | | : | | : | | : | |
118 glmynttrialphlltskglyvgmslagrvny-----inaavygakfigvnaifsetlge 172
: | : | : | : | | : | | : | | : | | : | |

Db 186 F-KGEIAVNALMPKTA-----LHTTAMDMLGGPGIE--SQCXY--DIIDAAYS 230
: | : | : | | | | | | | | | | | | | |
173 vergyrrvviepgtldelrghichtackem-----yegrisqirklgaqdlaaeavrya 227
: | : | : | | | | | | | | | | | | | |

OY 231 i 231
DB 228 v 228

RESULT 4
R77866 standard; Protein; 247 AA.
XX R77866;
XX
DT 13-NOV-1995 (first entry)
```

xx	S. clavuligerus ORF9 product.
DE	
xx	Clavulinic acid; clavulanate; antibiotic; beta-lactamase-inhibitor.
xx	
xx	Streptomycetes clavuligerus.
OS	
xx	CA2108113-A.
PN	
xx	
xx	09-APR-1995.
PD	
xx	
PF	08-OCT-1993; 93CA-2108113.
xx	
PR	08-OCT-1993; 93CA-2108113.
XX	
PA	(UYAL-) UNITV ALBERTA.
PI	
Aidoo KA,	Jensen SE, Paradkar AS;
DR	WPI; 1995-207301/28.
N-PSDB:	Q91580.
CC	
PT	Clavulinic acid biosynthesis enzymes and corresp. DNA - useful for
PT	biosynthesis of the antibiotic in Streptomycetes hosts which do not
PT	naturally produce clavulanate
PS	Claim 31; Fig.18; 41pp; English.
xx	
CC	A 15 kb fragment S. clavuligerus NRRL 5741 genomic DNA (Q91580),
CC	extending downstream from pcbC, included 10 ORFs encoding the
CC	enzymes required for clavulanate biosynthesis. ORF8 (oriented
CC	in the opposite direction to pcbC) encoded a protein (R77866)
CC	that showed weak similarity to ribitol 5 P04 dehydrogenase-type
CC	enzymes.
xx	
SQ	Sequence 247 AA:
	Query Match 10.5%; Score 224; DB 16; Length 247;
	Best Local Similarity 29.5%; Pred.No.1,4e-13;
	Matches 71; Conservative 40; Mismatches 96; Indels 34; Gaps
OY	8 LAGCTVFITGASRGICKATLAKADGANIVIAKTAOPHPKLIGTIYAAEEIEAVSGK 67
DB	5 Iqgkvalltggassjgeataraaeagaavaalarrve-----klralgdeltaagk 57
OY	68 ALPCIVDVADEEQISAAYKAIRKREGIDILVNNSAISLTMTLDTPTRDLMMNVNR 127
DB	58 vhwleldvdrdgvdvaasvtvealgydlilvnngimlllgpvedaiddctwrmdtl 117
OY	128 GTYLASKACIPYLKRSK--VAHILNISPPLNLNPWFROHCAYTTIARYGSMYVLGMAE 185
DB	118 glumymtraiphlrlskglvvgmslaagrwnv-----inaavyqatkfgvnaifsetlrg 172
OY	186 F-KGEIYNVALMPKTA-----IHTAAMDMLGGPGIE---SQCARKV---DIADAAYS 230
DB	173 vvergryvviviepgttidelrghitchtakem----yegrtisqrklkgaglaeevrya 227
OY	231 I 231
DB	228 v 228
RESULT	5
Y54422	Y54422 standard; Protein; 248 AA.
AC	Y54422;
DT	06-APR-2000 (first entry)
DE	Amino acid sequence of a beta-ketoacyl-ACP reductase protein.

DB 170 etlaIeyapkg-irvnnlgp-gaint 193

RESULT 7

ID Y54424 standard; Protein: 261 AA.

AC Y54424;

DT 06-APR-2000 (first entry)

DE Amino acid sequence of the glucose dehydrogenase enzyme.

KM Glucose dehydrogenase gene: Type II fatty acid synthetase;

KW stereoselectivity; 4-chloroacetoacetic acid ester;

KM (S)-4-halo-3-hydroxybutyric acid ester; 4-halo-acetoacetic acid ester;

KM beta-ketoacyl-acyl carrier protein reductase; acetoacetyl-CoA reductase;

KM polybeta-hydroxy fatty acid biosynthesis; optically active;

KM 4-haol-3-hydroxybutyric acid ester.

OS Bacillus subtilis.

PN EP955375-A2.

PD 10-NOV-1999.

PF 10-MAY-1999; 99EP-0109403.

PR 08-MAY-1998; 98JP-0126507.

PR 21-OCT-1998; 98JP-0300178.

PR 05-APR-1999; 99JP-0098205.

PA (DAIL) DAICEL CHEM.IND LTD.

PI Yamamoto H;

DR WPI: 2000-118183/11.

DR N-PSDB: 245757.

PT Producing optically pure (S)-4-halo-3-hydroxybutyric acid ester -

PS Example 18; Page 27-28; 34pp; English.

XX The present sequence represents the glucose dehydrogenase protein of

CC Bacillus subtilis. The glucose dehydrogenase gene was cloned, in

CC the course of the invention, to regenerate the reduced nicotinamide

CC adenine dinucleotide phosphate. The specification describes a

CC method for producing a (S)-4-halo-3-hydroxybutyric acid ester.

CC The method comprises asymmetrically reducing 4-halo-acetoacetic

CC reductase constituting Type II fatty acid synthase, or acetoacetyl-CoA

CC reductase constituting the polybeta-hydroxy fatty acid biosynthesis

CC system. The novel method is used to produce optically active

CC 4-haol-3-hydroxybutyric acid ester, with a high purity.

CC Sequence 261 AA:

XX

XX

XX

OY 180 LGMAEEF--KGETAVNALMPKTAIHT 203

DB 170 etlaIeyapkg-irvnnlgp-gaint 193

RESULT 8

ID Y30338 standard; Protein: 900 AA.

AC Y30338;

DT 15-NOV-1999 (first entry)

DE A multifunctional enzyme designated Fox2.

KM Multifunctional enzyme: Fox2; plant metabolism; transgenic plant;

KM fatty acid oxidation enzyme; polyhydroxyalkanoate; oil composition;

KM seed production; plant biomass.

OS Saccharomyces sp.

PN WO945122-A1.

PD 10-SEP-1999.

PF 05-MAR-1999; 99WO-US04999.

PR 06-MAR-1998; 98US-0077107.

PA (META-) METABOLIX INC.

PI Boynton L, Huisman GW, Moloney M, Patterson N, Peoples OP;

PI Snell K;

DR WPI: 1999-540850/45.

DR N-PSDB: 210369.

PT Modifying fatty acid metabolism in plants, useful for increasing

PT biomass and producing specific polymers in seeds

PS Disclosure; Page 52-56; 79pp; English.

XX The present sequence represents a multifunctional enzyme designated Fox2.

CC Such multifunctional enzymes may be used in the method of the invention.

CC The specification describes a method for manipulating the metabolism of

CC a plant, and comprises expressing a heterologous gene encoding fatty acid

CC oxidation enzymes in the cytosol or plastids other than the peroxisomes,

CC glyoxisomes or mitochondria of the plant. The method may be used to

CC enhance the biological production of polyhydroxyalkanoates or novel oil

CC compositions in a transgenic plant. Plants which may be used to produce

CC these compounds in this way include Brassicas, maize, soybean,

CC cottonseed, sunflower, palm, coconut, safflower, peanut, mustards, flax,

CC tobacco and alfalfa. The method may also be used to prevent or suppress

CC seed production and therefore increase the production of biomass (leaves,

CC stems, stalks) by plants.

XX Sequence 900 AA:

Query Match 10.1%; Score 217; DB 21; Length 261;
Best Local Similarity 31.6%; Pred. No. 7.2e-13;
Matches 65; Conservative 42; Mismatches 83; Indels 16; Gaps 7;

OY 1 MPTNTRLACGIVFTTGASRGIGKAIALKAARDGANIVIAAKTAQPHPKLLGTITTAEE 60
DB 1 mYPD--lkgkvaaltgaasgigkamaItfgkqkvvnyvsnkqdpn-----evkee 51
OY 61 IAVAGKALPCITVDVDEQOISAWEKAIKKFGIDILVNNASATSLTNTLDPTRKRLD 120
DB 52 vIkaggeavvvgdvtkedvknivqtaIkefgtIdImnnagIenpvshempIkdwk 111
OY 121 MNMNTRTGYTLASKACIPYLKRSKV-AHLINISPLNLNPFVKQHCATYIAKYGSMYV 179
DB 112 vgtlnltgtfsgsrealkyfvendIkgnvInmsvnevIpwplfvh--yaaakgglkInt 169

Query Match 9.9%; Score 211; DB 20; Length 900;
Best Local Similarity 26.0%; Pred. No. 1.8e-11;
Matches 109; Conservative 50; Mismatches 147; Indels 114; Gaps 21;

OY 13 VFTGASRGIGKAIALKAARDGANIVIAAKTAQPHPKLLGTI-----YTAAEEIE 62
DB 12 vltgsgggslgkvyalaIaysrgakvv-----ndlgltlgsgsgnskaadIvvdelk 63
OY 63 AVGKALPCITVDVDEQOISAWEKAIKKFGIDILVNNASATSLTNTLDPTRKRLDLM 122
DB 64 kagIavanydsvneenge--kIletakIefgrvdvIInnagIldvdfakmterefasv 121
OY 123 NNTNRTGYTLASKACIPYLKRSKVVAHLINISPLNLN--NFWMKQHCATYIAKYGSMYV 180

```

Db      122 dvhltygylklsraawpymrsgkfgrlntaspaglfgn--fgq-aynsaakmg1-----v 173
      181 GMAEEF-----KGEIANNALWPKTAIHTAAMDMLGSGPGIESQCRKVDIADAAV----- 229
      Db      174 glaeclakegakaylnvnslap--larsrmtenvlpplhlqldpkektvplvlyltheest 231
      QY      230 ----SIFQPKRSFTGN-----FVIDENILKEGCIEN-----FDVYAIKPGHPLOPD 271
      Db      232 kvansifelaagffgqlrwerssqglfnppdbktytpaellnkwlkeldydrkfnkcdhp 291
      QY      272 FFLDEYEAWSKVESGAVPEFEKEKLOLOPKRSGAVEETFRIVADSLSDDVYKATQA 331
      Db      292 yqlsdynldlt-----kakk1--pneqsgsvk-----lk-slcnkvvvvt-- 328
      QY      332 IYLFELSGEDSG-----TWFL-----DLKSKG-----NVGGEPS-----DQADV 368
      Db      329 -----gagggdglksnawfarygakvvnvndlkdfsvveelnklygegtalpdshdv 381

```

RESULT 9

R89323
ID R89323 standard; Protein; 315 AA.

AC R89323;

DT 13-APR-1996 (first entry)

DE Rape leaf beta-ketoacyl-ACP-ketoreductase.

XX Rape; leaf; beta-ketoacyl-ACP-ketoreductase; beta-ketoreductase;
KW plasmid pJR16.2; CDNA library; Escherichia coli; vector;
KW plasmid; stroma; transit peptide; cassette; antisense; oilseed;
KW transgenic plant; crop improvement; lipid; metabolic engineering;
KM polymer; rapeseed oil.

OS Brassica napus.

XX Key Location/Qualifiers
FT Peptide 1..55 /note= "Transit peptide"

XX W09602652-A2.

XX PD 01-FEB-1996.

XX PF 17-JUL-1995; 95WO-G801678.

XX PR 20-JUL-1994; 94GB-0014622.

XX PA (ZENE) ZENECA LTD.

XX PI Chase D, Elborough K, Fentem PA, Slabas AR, White A;

XX DR WPI: 1996-105914/11.

XX DR N-PSDB; Q99305.

XX New isolated rape beta-ketoreductase DNA - used to develop plants
PT with lower or higher oil contents or with altered oil compn.

XX PS Claim 2; Page 16; 29pp; English.

XX The sequence corresponds to a rape leaf beta-ketoreductase encoded
CC by a CDNA insert in plasmid pJR16.2 in Escherichia coli XL1-Blue.
CC A plasmid stroma targeting transit peptide is present. DNA
CC encoding the protein may be inserted in a vector or expression
CC cassette in sense or antisense orientation for expression in an
CC oilseed plant, e.g. for production of transgenic rape plants with
CC low or modified oil content, diversion of metabolism to alternative
CC storage compounds, e.g. starch, protein or engineered polymers, or
CC production of plants with enhanced oil content. The DNA may also
CC be used as a probe to obtain similar genes from other plants. The
CC transit peptide may be used to direct other proteins to leaf

CC plastids.
XX SQ Sequence 315 AA;

Query Match 9.78; Score 206.5; DB 17; Length 315;
Best Local Similarity 27.48; Pred. No. 1e-11;
Matches 69; Conservative 45; Mismatches 95; Indels 43; Gaps 7;

```

QY      13 VETIGASRGICKATVALNKAQGANIVI---AAKTAOPHPKLGTITTAABEIAVAGKA 68
      Db      75 vvvvgasrgigkatalalsgkagckvlnyarsakee-----evskqleaygga 124
      QY      69 LPCIVDNDEQOIAAVERAKIRKEGIDILVNNAISAISLTNTLTPPKRLDLMNVNTRG 128
      Db      125 ltfgdvskesadveammktaldawgtldvvnngitrdtlllmkksqdevldnlgt 184
      QY      129 TYLSKACIPYLLKSSKVAHILNISPLNLNFWKRONCAVITATYGSMTYVLGMAEEFKG 188
      Db      185 vflctgaatkimkkkrxgrlntasvyl--lgnlgqanyaaakag---vlgfsteaar 238
      QY      189 E-----IAVNALMP-----KTAIHTAAMDMLGSGPGIESQCRKVDIADA 227
      Db      239 egaarnlnvnnvcpgfllasdmktaklgedmekkllgtlplrgygp--edvaglvefials 296
      QY      228 AYSIFQPKRSFT 239
      Db      297 paasyiltgaf 308

```

RESULT 10

R89322
ID R89322 standard; Protein; 315 AA.

AC R89322;

DT 13-APR-1996 (first entry)

DE Rape seed beta-ketoacyl-ACP-ketoreductase.

XX Rape; seed; beta-ketoacyl-ACP-ketoreductase; beta-ketoreductase;
KW plasmid pJRS10.1; CDNA library; embryo; Escherichia coli; vector;
KW plasmid; stroma; transit peptide; cassette; antisense; oilseed;
KW transgenic plant; crop improvement; lipid; metabolic engineering;
KM polymer; rapeseed oil.

OS Brassica napus.

XX Key Location/Qualifiers
FT Peptide 1..56 /note= "Transit peptide"

XX W09602652-A2.

XX PD 01-FEB-1996.

XX PF 17-JUL-1995; 95WO-G801678.

XX PR 20-JUL-1994; 94GB-0014622.

XX PA (ZENE) ZENECA LTD.

XX PI Chase D, Elborough K, Fentem PA, Slabas AR, White A;

XX DR WPI: 1996-105914/11.

XX DR N-PSDB; Q99304.

XX New isolated rape beta-ketoreductase DNA - used to develop plants
PT with lower or higher oil contents or with altered oil compn.

XX PS Claim 1; Page 15; 29pp; English.

XX The sequence corresponds to a rape seed beta-ketoreductase encoded

Query Match	Best Local Similarity	Matches	Conservative	Score	DB	Length	Indels	Gaps
10	9.5%	72	53	202.5	20	313	39	12
7	25.4%	72	53	202.5	20	313	39	12
60	25.4%	72	53	202.5	20	313	39	12
122	25.4%	72	53	202.5	20	313	39	12
120	25.4%	72	53	202.5	20	313	39	12
182	25.4%	72	53	202.5	20	313	39	12
177	25.4%	72	53	202.5	20	313	39	12

```

Oy      229  YSIFQPKK--SFTGNVVIDENILKEGGIEFPVNAIKRGHLO 269
      :: | | | | | | | | | | | | | | | | | | | | | |
Db      237  valatdpnlislsqkvlpacdlaaryldvrd-----grpvq 273

RESULT  13
ID       R27757
R27757  R27757 standard; protein; 261 AA.
XX
XX      R27757;
XX
XX      11-MAR-1993 (first entry)
XX
XX      Glucose dehydrogenase.
XX
XX      GDH; mutant; recombinant; mass production; tetramer; thermostable.
XX
XX      Bacillus megaterium.
XX
XX      Key      Location/Qualifiers
XX      FET      MISC-difference 96      /note= "site of Glu->Val or Lys mutation"
XX      FET      MISC-difference 252     /note= "site of Glu->Leu mutation"
XX      FET      MISC-difference 253     /note= "site of Tyr->Glu mutation"
XX      FET      MISC-difference 22      /label= Ser, Ala
XX      FET      MISC-difference 43      /label= Asp, Glu
XX      FET      MISC-difference 79      /label= Ala, Ser
XX      FET      MISC-difference 95      /label= Leu, Met
XX      FET      MISC-difference 95      /label= Leu, Met
XX
XX      JP04258293-A.
XX
XX      14-SEP-1992.
XX
XX      13-FEB-1991; 91JP-0106927.
XX
XX      13-FEB-1991; 91JP-0106927.
XX
XX      (AMAN ) AMANO PHARM KK.
XX
XX      WPI; 1992-354684/43.
XX
XX      Glucose dehydrogenase prepn. using transformed recombinant DNA
XX      from Bacillus megaterium - has specified transformations giving
XX      glucose dehydrogenase-expressing vector, introduced into E.coli.
XX      for culture
XX
XX      Claim 1; Page 8; 12pp; Japanese.
XX
XX      The glucose dehydrogenase enzyme is encoded by a recombinant DNA
XX      clone from Bacillus megaterium. The DNA sequence may be mutated by
XX      site directed mutagenesis to introduce mutations to the protein
XX      sequence (see feature table). The DNA may be used to transform E.
XX      coli cells, and transformants may be cultured to mass produce GDH.
XX      The mutant GDH is not influenced by ion strength, exists as a
XX      tetramer and is thermostable.
XX
XX      Sequence      261 AA:

Query Match      9.3%; Score 199.5; DB 13; Length 261;
Best Local Similarity 29.1%; Pred. No. 3.6e-11;
Matches 58; Conservative 44; Mismatches 84; Indels 13; Gaps

Oy      8  LAGCTVTTASNGICAKAIALKAADGANIYAIAKTAQPHFKLLGTYTAAETIEAVGSK 67
      | | | | | | | | | | | | | | | | | | | | | |
Db      5  legvvvvltsscgllkxmalrfatekakvvyvnyksxexaan-----svleelkkvvgg 58

```

[illegible]

```

Db      119 gafisreaiakyfvendikgvtimmsvhek1pwlfnh-yaaskgymkxmwen1aley 176
OY      187 --KGEIAVNALMPKTA1HT 203
Db      177 apkg-1rvnnigp-gaint 193

RESULT 15
Y55986
ID      Y55986 standard; Protein; 251 AA.
AC      Y55986;
XX      15-MAR-2000 (first entry)
DT      S.clavuligerus cladh protein.
DE      S.clavuligerus cladh protein.
XX      Operon: cladh; clar: biosynthetic pathway: clavulanic acid; enzyme:
KW      clavulanate-9-aldehyde reductase; transcriptional regulator;
KM      gene expression.
XX      Streptomyces clavuligerus.
OS      Streptomyces clavuligerus.
XX      ES2131001-A1.
PN      01-JUL-1999.
PD      16-JUN-1997; 97ES-0001305.
PF      16-JUN-1997; 97ES-0001305.
XX      PR      16-JUN-1997; 97ES-0001305.
XX      PA      (ANTI ) ANTIBIOTICOS SAU.
XX      WPI; 2000-026016/03.
DR      N-PSDB; 230700.
XX      PT      Improving production of clavulanic acid by Streptomyces - by
PT      super-expression of clar gene.
XX      Example 1; Page 8-9; 23pp; Spanish.
XX      This sequence represents the cladh protein which is a putative
CC      clavulanate-9-aldehyde reductase used in the biosynthetic pathway for
CC      clavulanic acid in Streptomyces clavuligerus. The invention relates to
CC      methods of improving production of clavulanic acid by Streptomyces by
CC      super-expression of clar gene. The clar gene is characterised in that
CC      it is localised in the gene group encoding genes for biosynthesis of
CC      clavulanic acid.
XX      Sequence 251 AA;
SQ

Query Match 9.3%; Score 198; DB 21; Length 251;
Best Local Similarity 27.8%; Pred.No.4.7e-11;
Matches 69; Conservative 40; Mismatches 103; Indels 36; Gaps 9;

OY      1 MDPNNGRLAGCTVFVTGASRGIGKAIATLKAADGANIVIAAKTAOPHPKLTGITYTAAEE 60
Db      4 mmpsa--1ggkkaalltggelgrrtaratalapagaavaataarve-----k1a1gde 54
OY      61 IEAVGKALPCTIVDVRDEQOISAAVEKA1KKRGGIDILVNNASAI5LNTLDTPTKRRLD 120
Db      55 ltaagakhvleldvdrdgvaavaastveaalgyldilvnaaglm1lgpredadctdwr 114
OY      121 MANNVTREGTYLASKACIPILKRSK--VAHILNISPPILNLPWFQKHCAITYIAKGMSHY 178
Db      115 midtl1l1mymtcral1ph1l1rskgtlvvgmsia-----grvtvnaavygatk1gvnaf 169
OY      179 VLGMAEF-KGEIAVNALMPKTA-----1HTAMDLGPGIE--SOCRKY--DI 223
Db      170 setvqetvgrvrvv1vlepytdtel1gth1htatkem-----yegrlsq1rk1qagdi 224

```

Oy	224	IADAYSI	231
		1::	
Db	225	aeavryav	232

Search completed: June 11, 2001, 16:23:10
Job time: 140 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 11, 2001, 16:23:15 ; Search time 29.17 Seconds
(without alignments)
1679.565 Million cell updates/sec

Title: US-09-464-039-7

Perfect score: 2139

Sequence: 1 MHPNTRGLAGCTVFITGASR.....GNMALAIKIKELKLNOMNARL 418

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL.15:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.podent:*
13: sp.rodent:*
14: sp.unclassified:*
15: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1179.5	55.1	412	5	09VB10	09VB10 drosophila
2	1034	48.3	938	5	009979	009979 caenorhabdi
3	891.5	41.7	293	5	018639	018639 caenorhabdi
4	665	31.1	282	2	005842	005842 mycobacteri
5	351	16.4	441	5	09NKM1	09NKM1 dictyosteli
6	302.5	14.1	1015	3	09UVH9	09UVH9 glomus moss
7	286	13.4	436	5	021481	021481 caenorhabdi
8	267	12.5	248	2	044326	044326 agrobacteri
9	245.5	11.5	735	13	042484	042484 gallus gall
10	244	11.0	240	1	09UY54	09UY54 pyrococcus
11	235	10.9	246	2	09KA03	09KA03 bacillus ha
12	233.5	10.6	255	2	P95286	P95286 mycobacteri
13	227	10.6	253	2	09K3Y7	09K3Y7 streptomyce
14	227	10.6	679	2	09XAB2	09XAB2 streptomyce
15	224.5	10.5	301	2	053398	053398 mycobacteri
16	224	10.3	247	2	09LCV7	09LCV7 streptomyce
17	221	10.2	248	2	P71824	P71824 mycobacteri
18	218.5	10.2	248	2	09RH22	09RH22 zymomonas m
19	218	10.2	251	5	09U112	09U112 drosophila

RESULT	1	ALIGNMENTS
09VB10	PRELIMINARY;	PRT: 412 AA.
AC 09VB10:		
DT 01-MAY-2000 (TREMBLrel. 13, Created)		
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)		
DE CG5590 PROTEIN.		
GN CG5590.		
OS Drosophila melanogaster (Fruit fly).		
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC Ephydroidea; Drosophilidae; Drosophila.		
OX NCBI_TaxID=1227;		
RP [1]		
RC SEQUENCE FROM N.A.		
RX STRAIN=BERKELEY;		
RA MEDLINE=20196006; PubMed=10731132;		
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,		
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,		
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA Abul J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,		
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolintsov S.,		
RA Borokva D., Botchan M.R., Bouck J., Brostein P., Brothier P.,		
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,		
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA de Palos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA Dudzik K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,		
RA Foster C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,		
RA Hostin D., Houston K.A., Howard T.J., Wei M.-H., Ibbegam C.,		
RA Jattali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,		
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,		
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,		
P95273 mycobacteri		
034187 halobacteri		
09T87 streptomyce		
09ZB95 rhodococcus		
007882 staphylococ		
09K636 bacillus ha		
09R87 klebsiella		
023618 caenorhabdi		
091269 streptomyce		
091061 streptomyce		
09W5C0 drosophila		
016969 caenorhabdi		
09X4W7 pseudomonas		
09VX10 drosophila		
09S2E4 streptomyce		
09N538 caenorhabdi		
09VNF3 drosophila		
P70540 ratius norv		
056840 xanthobacte		
09PCQ2 xyella fas		
069366 rhodococcus		
091b50 bacillus th		
091b50 caenorhabdi		
09VNF2 drosophila		
P70523 ratius norv		
09K5C3 burkholderi		

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkimos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang Q., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003762; AAF56735.1; -
 DR HSSP: P47227; 1BD8.
 DR FLYBASE: FBgn0039537; CG5590.
 DR INTERPRO: IPR002198; -
 DR INTERPRO: IPR002347; -
 DR INTERPRO: IPR003033; -
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF02036; SCP2; 1.
 DR PRINTS: PRO0081; GDHRDH.
 SO SEQUENCE 412 AA; 44354 MW; 2136F5F07FCAE0B CRC64;

Query Match 55.1%; Score 1179.5; DB 5; Length 412;
 Best Local Similarity 56.4%; Pred. No. 3.9e-75;
 Matches 238; Conservative 62; Mismatches 103; Indels 19; Gaps 5;

QY 4 MNGRLAGCVFTTGSARGIGKALAKAARDGANIYAATAOPHRLTGTIYAAEEIA 63
 DB 3 MNGRLAGRLFTTGSARGIGKELALKAARDGANIYAATAOPHRLTGTIYAAEEIEK 62
 QY 64 VGGKALPCIVDRDEQOISAAYEKAIRKFGIDILVNNASISLTNTLPTPKRLDMMN 123
 DB 63 AGCKAPCVDRDEQOVSAYEAAVAKGIDIVNNASISLTNTPTDKRYDLMHN 122
 QY 124 VNRGTLYLAKACIPYLKSKVAHILNISPLNLNPFVKQHCATYIAYGSMYVLGMA 183
 DB 123 INVRGFLVLSKCLPYLKSNNAHILNISPLSMKMGFPVAVYAYGSMCYLGNMA 182
 QY 184 EEFGKE-IVNMLMPKTAHTAAMDMLGGPGIESQCRKDIADAYSYF-OKPKSFTGN 241
 DB 183 AERKDEGISVNLMPKTAHTAHTAEMLTGPDSAKSKRPEIMADAYALTLREPRSTQ 242
 QY 242 FVIDENILKEEGIEENEDVYAI--KPGHPLOPDFLDEYPEAVSKVSTGAVPEFEKEL 299
 DB 243 FVYDDEVLSSAGITDLEVAACFRENADKLMVDFVEKCAPVENEAAADDA----- 294
 QY 300 QIQPKRSGA---VEETPRIVKDSLDVYKATQATYLFELSGEDGTFPLDKSGGVN 356
 DB 295 ---APASGADVAKIQPLFKIESLSPLEIVSKTQAFQFNISGAEQGTFLDKNGSGSC 350
 QY 357 GGEPSDQADVMSMTDDFVFMFSGKLKPTAFMFGSKLKIGNNALAIKLEKLNQMA 416
 DB 351 GAGTPAAPDALTLMNSKNFPMFSGKLAAAPYATMGTKLISDFOKALKLEKLMALKS 410
 QY 417 RL 418
 DB 411 KL 412

RESULT 2
 ID 009979 PRELIMINARY: PRT: 938 AA.
 AC 009979:
 DT 01-JUN-1998 (TREMBlrel. 06, Created)
 DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE HYPOTHETICAL 105.9 KDA PROTEIN C17G10.8 IN CHROMOSOME II.
 GN C17G10.8.
 OS *Caenorhabditis elegans*.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; *Caenorhabditis*.
 NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Johnson D.;
 RN Submitted (Jun-1995) to the EMBL/Genbank/DBJ databases.
 RC STRAIN-BRISTOL N2;
 RA Waterston R.;
 RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: STRONG IN THE C-TERMINUS TO C.ELEGANS C45B11.3 AND
 CC MYOACTERIUUM TUBERCULOSIS MTCY07D11.02C.
 DR EMBL: U28739; AAB93456.1; -
 DR MORMEP: C17G10.8; CE18681.
 DR INTERPRO: IPR002198; -
 DR INTERPRO: IPR003033; -
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF02036; SCP2; 1.
 KW Hypothetical protein.
 SO SEQUENCE 938 AA; 105896 MW; D64A09FBBC76C775 CRC64;

Query Match 48.3%; Score 1034; DB 5; Length 938;
 Best Local Similarity 54.6%; Pred. No. 2.4e-64;
 Matches 230; Conservative 52; Mismatches 111; Indels 28; Gaps 8;

QY 7 RLACGCVFTTGSARGIGKALAKAARDGANIYAATAOPHRLTGTIYAAEEIAYVG 66
 DB 526 KEVGRVLTITGSARGIGKELALKAARDGANIYAATAOPHRLTGTIYAAEEIEKAG 585
 QY 67 KALPCIVDRDEQOISAAYEKAIRKFGIDILVNNASISLTNTLPTPKRLDMMNVT 126
 DB 586 KALPCIVDRDEASVAYEAAVAKGIDILVNNASISLTNTPTDKRYDLMHN 125
 QY 127 RGTLYLAKACIPYLKSKVAHILNISPLNLNPFVKQHCATYIAYGSMYVLGMA 186
 DB 646 RGTFLMTKCLPYLKSNNAHILNISPLMETRMFANVAYATMAKYGSMCYLGNMA 705
 QY 187 KGE-IVNMLMPKTAHTAAMDMLGGPGIESQCRKDIADAYSYFQK-PRKSTGNFY 244
 DB 706 RRGIVNMLMPKTAHTAHTAEMLTGPDSAKSKRPEIMADAYALTLREPRSTQ 242
 QY 245 DENILKEEGIEENEDVYAIKPGHPLOPDFLDEYPEAVSKVSTGAVPEFEKELQDPK 304
 DB 766 DEDILKAEQVTDPRVACVPDAPLMPDFI---PACTYDHKFSGA-----QIGKK 813
 QY 305 PR---SGAVEE---TFRIYKDSLDVYKATQATYLFELSGEDGT-----WFLDKSK 352
 DB 814 NKTHEAGVVEEIKQIFTSAKRLNLNDIVKKTGEVYFLL--KDPTTKSERITITDLKNG 871
 QY 353 GGNVGEPSDQADVMSMTDDFVFMFSGKLKPTAFMFGSKLKIGNNALAIKLEKLN 412
 DB 872 EGALTPKKAAGKADVFTLAPENHAPLFTGKLRPTALTMTKLIQSDGPMGAKLESLLR 931
 QY 413 Q 413
 DB 932 K 932

RESULT 3
 ID 018639 PRELIMINARY: PRT: 293 AA.
 AC 018639:
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
 DE C45B11.3 PROTEIN.
 GN C45B11.3.
 OS *Caenorhabditis elegans*.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

"2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*.";
 RL Nature 368:32-38(1994).
 [2]
 RA Du Z.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).
 DR EMBL: U41544; AAA83184.1; -.
 DR HSSP: P25529; IPMC.
 DR INTERPRO: IPR002198; -.
 DR INTERPRO: IPR003033; -.
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF02036; SCP2; 1.
 DR PROSITE: PS00061; ADH_SHORT; UNKNOWN_1.
 DR Oxidoreductase.
 KW Plasmid.
 SO SEQUENCE 436 AA; 46262 MW; 2DBD94562A09BE55 CRC64;
 Query Match 13.4%; Score 286; DB 5; Length 436;
 Best Local Similarity 25.8%; Pred. No. 3e-12;
 Matches 120; Conservative 73; Mismatches 176; Indels 96; Gaps 17;
 QY 7 RLACCTVITGASRGICAKALKAADGANIVIAKTAQPHKLLGTYTAA-----EE 60
 Db 4 RFDGKVAIVTGAGGGLGTALBELAKRGCKVYVNDLGGDRH---GTSSSSSMADKVVQ 59
 QY 61 IEAVGKALPCTIVYRDEQOISAAVEKAIKFGGIDILVNNASISLTNTLDTPTKRLD 120
 Db 60 IKSAGGAVANYDYSEFGDKI--VKTAIDNFGRIYINNAGILRDVYFLKMTLMDL 116
 QY 121 MANNTRGTYLASKACIPYLKSKVAHILNISPPLNLPWFQKHCAYTIKYGSMYVL 180
 Db 117 IFKYNKGAIVATKAAWYMDQKGRIVYSSNAGVGNQGAN--YAAKSAKALIGLSN 174
 QY 181 GMAEE-FKGEIAVNALMP-----KTAIHNAAMD----- 209
 Db 175 SLAEGAKYNIILANTLVYTAGSRLETVMQNLVDALRDPYVPLVTVYVHDSFEESKV 234
 QY 210 --GGPG----IESQCRKDIIDAAVSTFOCKPSTGTFVIDENT--LKEGIEFQVYA 261
 Db 235 FEAGAGWGTIOYKSKGIVSHASADDIAKNWSTTNMNGAEYIGTTESARLVST-- 292
 QY 262 IKPGHPIQPDFLDEY-----PEAVSKVESTGAVPEKEEKLQPKPRGAVEETPRI 316
 Db 293 -----LEHEHSSGSSSSSSGASGGAFFS-----NIRSSAL--FOE 327
 QY 317 VKDSLSD--VVKATQAIYLFELSGEDG---GTWFLDKSKGNGVYGE--PSDAQDV 368
 Db 328 MADGVKADPVAVKTLKSLIVYIT--DGKNELGKFTLDFKASPSVYIGDVKNCKANAT 385
 QY 369 MSMTTDDVVKMFSGKLTAFMSGKLIKIKGNMALATLEKLMNG 413
 Db 386 VTVADSDVFVDIAAGKLNQAKAFMSGKLVKGNVMLQLQTVLEK 430
 RESULT 8
 ID 044326 PRELIMINARY; PRT; 248 AA.
 AC 044326;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE MOC PROTEIN.
 GN MOC.
 OS Agrobacterium tumefaciens.
 OG Plasmid pT15955.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Agrobacterium.
 OX NCBI_TaxID=362;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=15955;
 RX MEDLINE=96236046; PubMed=8655509;
 RA Kim K.S., Farrand S.K.;
 RT "Plasmid encoded genes responsible for catabolism of the crown gall
 RT opine mannopine by Agrobacterium tumefaciens are homologs of the r-
 RT region genes responsible for synthesis of this opine by the plant
 RT tumor.";
 RL J. Bacteriol. 178:3275-3284(1996).
 DR EMBL: AF242881; AAB07783.1; -.
 DR INTERPRO: IPR002198; -.
 DR INTERPRO: IPR002347; -.
 DR PFAM: PF00106; adh_short; 1.
 DR PRINTS: PR00080; SDRFAMILY.
 DR PRINTS: PR00081; GDRHD.
 KW Plasmid.
 SO SEQUENCE 248 AA; 26810 MW; 10B69239CDEBB68D CRC64;
 Query Match 12.5%; Score 267; DB 2; Length 248;
 Best Local Similarity 34.6%; Pred. No. 2.7e-11;
 Matches 71; Conservative 41; Mismatches 71; Indels 22; Gaps 8;
 QY 7 RLACCTVITGASRGICAKALKAADGANIVIAKTAQPHKLLGTYTAAEEIYAVG 66
 Db 2 KLOGTAVITAGRGIGRATLELAKEGCNVLAIE-----LNEVEAAVEYRAIGS 54
 QY 67 KALPCTIVYRDEQOISAAVEKAIKFGGIDILVNNASISLTNTL-DTPTRKLDMMNVN 125
 Db 55 EALALRTDVQKSEYDALAKAAFERFGAVDILVNA-GVAIHNTPIPKEDMDMMAIN 113
 QY 126 TRGYLASKACIPYLKSKVAHILNISPPLNLPWFQKHCAYTIKYGSMYVLGMAE- 184
 Db 114 LKGFELCTQAEFHQCDRHGHIIIVVS--RARKVASAKFAYVASKRGM---LGFTQT 167
 QY 185 -EFGK-ETAVNALMPKTAIHNAAMD 207
 Db 168 TDQEGIEFGVKA---TAVCPGAVD 188
 RESULT 9
 ID 042484 PRELIMINARY; PRT; 735 AA.
 AC 042484;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE 17-BETA-HYDROXYSTEROID DEHYDROGENASE TYPE IV.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97388669; PubMed=9245902;
 RA Kobayashi K., Kobayashi H., Ueda M., Honda Y.;
 RT "Expression of 17 beta-hydroxysteroid dehydrogenase type IV in chick
 RT retinal pigment epithelium.";
 RL Exp. Eye Res. 64:719-726(1997).
 CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES FAMILY (SDR).
 DR EMBL: U77911; AAC60249.1; -.
 DR HSSP: O12634; IYBV.
 DR INTERPRO: IPR002198; -.
 DR INTERPRO: IPR002539; -.
 DR INTERPRO: IPR003033; -.
 DR PFAM: PF00106; adh_short; 1.
 DR PFAM: PF01575; MOC_dehydratase; 1.
 DR PFAM: PF02036; SCP2; 1.


```
DB 2 LOGKTAIVTGSARGIGRATAMELARGANVYVYVNAKNEKAERYVA-----EIKELGV 54
OY 67 KALPCTIVDRDEQOISAAREKAIRKFGGIDILVNNASAI SLTNTLDTPTKRL-----DLM 121
DB 55 EAIATADVADESESOAVAKVETIDFGAVDILVNNAGI-----TRDNLFMKREEDMWAV 109
OY 122 MNVNTRGTYLASKACIPYLKSKVAHILNISPLNLNPVWFKQHCAYTTAKGSMYVLG 181
DB 110 IDTNLKGFFHCKSAVTRPMKORFRIINVS SVG-----ALGNAGQANVYAA 157
OY 182 MAEEKGEIAVNAIMPRTAHTAAMDMLGPGIESQCRKVDIADAAYSIFOKPKSFTGN 241
DB 158 KA-----GV 161
OY 242 EVIDENILKEGIEFNDYAIKRGHPLOPDPFLDEYEPNAVSKVSTGAVPE-FKEEKLO 300
DB 162 IGLVITLRELALNRNITVNAAPG-----FTET-----TMTGELPBDVAKOMLG 205
OY 301 LQPKRSGAVETFRIVKDSLSDVVKAT-QAIVL 334
DB 206 QIPLARLQPEEVAKAVRFLASDDASLYLGCTIHV 240
RESULT 12
P95286 PRELIMINARY; PRT; 255 AA.
ID AC P95286;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
GN HYPOTHEETICAL 27.0 KDA PROTEIN.
OS RV1928C OR MTCY09F9.36.
OC Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=968295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Davlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL; Z84498; CAB06498.1; -.
DR HSSP; P50163; ZAE1.
DR TUBERCULIST; RV1928C; -.
DR INTERPRO; IPR002198; -.
DR INTERPRO; IPR002347; -.
DR INTERPRO; IPR003015; -.
DR PFAM; PF00106; adh_short; 1.
DR PFAM; PF00678; adh_short_C2; 1.
DR PRINTS; PR00080; SDRFAMILY.
DR PRINTS; PR00081; GDRDH.
DR PROSITE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
DR PROSITE; PS00061; ADH_SHORT; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 255 AA; 27031 MW; 454692E6FAA253FF CRC64;
```

```
DB 9 LHGRALITGASTGIGKRVALLAVENAGNAVIAARHDLAEL-----ADEITSGK 61
OY 68 ALPCTIVDRDEQOISAAREKAIRKFGGIDILVNNASAI SLTNTLDTPTKRLDLMNNVTR 127
DB 62 VPPVCCDVSOHQOVTSMIDQVTAELGIDIAVCNAGIITVTPMLDMPLEEFORLONTVNT 121
OY 128 GYLSKACIPYLK-----SKVAHILNISPLNLNPVWFKQHCAYTTAKGSMYVLG 176
DB 122 GVFLLQAQAANKAVKOGGCVIINTASMSGHITNV-----PQOVSHYCA--SKAAVT 171
OY 177 MYVLMAEEF-KGEIAVNALMP 197
DB 172 HLTKAMAVELAPHKIRVNSVSP 193
RESULT 13
ID AC O9K3Y7 PRELIMINARY; PRT; 253 AA.
ID O9K3Y7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DB PUTATIVE SHORT CHAIN OXIDOREDUCTASE.
GN 2SCG61.28C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RA Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Redenbach M., Kleser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinash H., Hopwood D.A.;
RA "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL359949; CAB95803.1; -.
SQ SEQUENCE 253 AA; 25937 MW; 608F0D5C1AE55A8A CRC64;
```

Query Match 10.6%; Score 227; DB 2; Length 253;
Best Local Similarity 31.1%; Pred. No. 1.8e-08;
Matches 64; Conservative 33; Mismatches 83; Indels 26; Gaps 5;

```
OY 15 ITGASRGIGKATALKAAKADGANIVY-----NAKRAQPHPKLGTITYAAEIEVAGKA 68
DB 11 VTGAARGIGATATLALALAEGRVAVITDDEACK-----DVEKITTAAGKA 57
OY 69 LPCIVDRDEQOISAAREKAIRKFGGIDILVNNASAI SLTNTLDTPTKRLDLMNNVTRG 128
DB 58 IAVGCDVSDQAQVEAAVAVARIAEELGAPITLVNNAAGVLDNLLFKMSYSDMDVTVMVHLRG 117
OY 129 TYLSKACIPYLKSKVAHILNISPLNLNPVWFKQHCAYTTAKGSMYVLGMAEEF-K 187
DB 118 AFLMTKACQKHVDAKFGFVNVLSSSSALGN--RGQVNYSAKAGLOGFTTTLAKELGK 174
OY 188 GEIAVNALMP--KTAHTAAMDMLG 210
DB 175 FGVTAANAAPGFATETMTKATADRVG 200
RESULT 14
O9XAB2
```

```
ID 09XAB2 PRELIMINARY; PRT; 679 AA.
AC 09XAB2;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE PUTATIVE OXIDOREDUCTASE.
GN SCF3A.03.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K., Harris D.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb streptomycos coelicolor A3(2) chromosome."
RN Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MEDLINE=97000351; PubMed=843436;
RA Redenbach M., Kleser H.M., Denapate D., Eicher A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RN Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL096837; CAB48890.1; -.
DR HSSP; P25529; 1AHH.
DR INTERPRO: IPR002198; -.
DR INTERPRO: IPR002347; -.
DR PFAM: PF00106; adh_short; 1.
DR PRINTS; PRO0080; SDRFAMILY.
DR PRINTS; PRO0081; GDHRDH.
SQ SEQUENCE 679 AA; 71668 MW; 3826C2684B73BBA4 CRC64;
```

```
Query Match 10.6%; Score 227; DB 2; Length 679;
Best Local Similarity 29.4%; Pred. No. 8.3e-08;
Matches 82; Conservative 41; Mismatches 114; Indels 42; Gaps 11;
```

```
QY 2 LPNTGRLACCTVITGASRGIGKALKAADGANIYIAKTAOPHPKLLGTITYTAAEI 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 412 MPRPKPLATRVALTGTGSGIGKALRLVDEGACVYVADLNAE-----NAAYA 461
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 62 EAVGG--KALPCIVDVNDEQOISAAVEKAIRKFGGIDILVNNASALSITNTLDTPTKR 119
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 462 EELGGDDKAAVAVDVYSEEOIAAFQAAALAFGVDLVVNNAGISISKPLLETSAKDMD 521
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 120 LMNVNVRGTYLASKACIPYLKSKV--AHILNTSPRLNLPVWF--KHCAVYTIKAYGMSM 177
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 522 LQHDIMARGSFVLSREARARVTAOELGDIYLA---SKNAVERAGPNNIAYISATKADQAH 578
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 178 YVLGMAEEFGE--IAVNALMPTAHTAAMDMLGPGIESOCKKYDIADAAYSIFQKP 235
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 579 QVRLAAEL--GEHGIRYNGVNPQGVVRSGL--FAGCGAKRA-----AVYGV---P 624
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 236 KSFTGNFVIDENILKEGIEENFDVYAIKPGHPLOPFFL 274
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 625 EEKIGEEFYAORTLLKRE-----VLPERVANA VFAL 654
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
RESULT 15
ID 053398 PRELIMINARY; PRT; 301 AA.
AC 053398;
DT 01-JUN-1998 (TREMblrel. 06, Created)
DT 01-JUN-1998 (TREMblrel. 06, Last sequence update)
```

```
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE OXIDOREDUCTASE.
GN RV1050 OR MTV017.03.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gao S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RL complete genome sequence."
RN Nature 393:537-544(1998).
DR EMBL; AL021897; CAAL17166.1; -.
DR HSSP; P14061; 1FDU.
DR TUBERCULIST; RV1050; -.
DR INTERPRO: IPR002198; -.
DR PFAM: PF00106; adh_short; 1.
SQ SEQUENCE 301 AA; 32497 MW; 3B608922D291031C CRC64;
```

```
Query Match 10.5%; Score 224.5; DB 2; Length 301;
Best Local Similarity 30.9%; Pred. No. 3.5e-08;
Matches 58; Conservative 40; Mismatches 77; Indels 13; Gaps 5;
```

```
QY 13 VFTGASRGIGKALKAADGANIYIAKTAOPHPKLLGTITYTAAEIRAVGKALPCI 72
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 11 VLTGASSGIGENTAKRFAREGAVVALAAR-----CALRRVAREIEAAGRAMVAP 63
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 73 VDVREDOQISAAVEKAIRKFGGIDILVNNASALSITNTLDTPTKRDL--MNVNTRGTY 130
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 64 LDVSSSEVRAMVADVGEGRIDVFNNA--GVSLGPDVAFETFLDTRMEIDVLTGV 122
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 131 LASKACIPYLKSKVNHILISPPNLNLPVWFKHCAVYTIKAYGMSYVLGMAEEFGE- 189
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 123 RVVREVLPIKKQDSGRIMWSSVYGRKA--FAFAGYSSAMHAIAGFSDALHQLRSGS 180
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 190 IAVNALMP 197
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 181 IAVSVIHP 188
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Search completed: June 11, 2001, 16:25:32
Job time: 137 sec
```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 11:40:23 ; Search time 1766 Seconds
(without alignments)
12668.826 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561
Sequence: 1 aggcagagaagtatgcaagca.....gsgmgiraswmwawtwrammc 2561

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 9623517 seqs, 4368049070 residues

Word size : 0

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estp11:*
73: em_estp12:*
74: em_estp13:*
75: em_estp14:*
76: em_estp15:*
77: em_estp16:*
78: em_estp17:*
79: em_estp18:*
80: em_estp19:*
81: em_estp110:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

190: em_gss_pln1:*
 191: em_gss_pln2:*
 192: em_gss_pro:*
 193: em_gss_r0d1:*
 194: em_gss_r0d2:*
 195: em_gss_r0d3:*
 196: em_gss_r0d4:*
 197: em_gss_r0d5:*
 198: em_gss_vrt1:*
 199: em_gss_vrt2:*
 200: em_gss_vrt3:*
 201: em_gss_vrt4:*
 202: em_gss_vrt5:*
 203: em_gss_vrt6:*
 204: em_gss_vrt7:*
 205: em_gss_vrt8:*
 206: em_gss_vrt9:*
 207: em_gss_vrt10:*
 208: em_gss_vrt11:*
 209: em_gss_vrt12:*
 210: em_gss_vrt13:*
 211: em_gss_vrt14:*
 212: em_gss_vrt15:*
 213: em_gss_vrt16:*
 214: em_gss_vrt17:*
 215: em_gss_vrt18:*
 216: em_gss_vrt19:*
 217: em_gss_vrt20:*
 218: em_gss_vrt21:*
 219: em_gss_vrt22:*
 220: em_gss_vrt23:*
 221: em_gss_vrt24:*
 222: em_gss_vrt25:*
 223: em_gss_vrt26:*
 224: em_gss_vrt27:*
 225: em_gss_vrt28:*
 226: em_gss_vrt29:*
 227: em_gss_vrt30:*
 228: em_gss_vrt31:*
 229: em_gss_vrt32:*
 230: em_gss_vrt33:*
 231: em_gss_vrt34:*
 232: em_gss_vrt35:*
 233: em_gss_vrt36:*
 234: em_gss_vrt37:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	896	35.0	977	105	AL514158
2	802	31.3	861	106	AL528484
3	584	22.8	602	141	BE893726
4	500	19.5	887	168	BF699109
5	491	19.2	620	111	AM152065
6	489	19.1	963	141	BE892079
7	442	17.3	541	122	AW966207
8	419	16.4	487	5	AA315514
9	418	16.3	692	113	AW299683
10	398	15.5	829	171	BF979524
11	393	15.3	462	5	AA305116
12	330	12.9	484	157	W04752
13	326	12.7	482	156	R56767
14	325	12.4	386	9	AA622988
15	317	11.6	340	159	Z19446
16	291	11.4	616	32	AA348726
17	291	11.4	616	32	AV714646
18	284	11.1	497	32	AV750182

C	19	275	10.7	447	156	R89462	YQ01b10.r1
	20	270	10.5	310	166	BE392496	BE392496.601309277
	21	262	10.2	620	31	AV701344	AV701344
	22	262	10.2	661	31	AV701163	AV701163
	23	258	10.1	289	5	AA337010	AA337010
	24	258	10.1	469	155	R24760	R24760
	25	257	10.0	369	169	BF818687	BF818687
C	26	245	9.6	442	104	AT961078	AT961078
	27	238	9.3	912	173	BG107324	BG107324
	28	217	8.5	340	169	BF818347	BF818347
C	29	215	8.4	437	149	BF509823	BF509823
C	30	188	7.3	275	159	225140	225140
	31	163	6.4	518	3	AA210776	AA210776
C	32	144	5.6	152	143	BF062984	BF062984
	33	144	5.6	319	156	R89463	R89463
	34	133	5.2	209	147	BF360596	BF360596
	35	129	5.0	408	147	BF351420	BF351420
	36	120	4.7	131	113	AM297867	AM297867
	37	118	4.6	396	31	AV652595	AV652595
	38	116	4.5	315	31	AV701438	AV701438
	39	113	4.4	543	31	AV703524	AV703524
	40	110	4.3	210	1	AA054853	AA054853
C	41	97	3.8	197	140	BE827340	BE827340
	42	95	3.7	308	6	AA345403	AA345403
	43	91	3.6	851	145	BE209703	BE209703
	44	85	3.3	187	140	BE85366	BE85366
	45	85	3.3	188	140	BE85273	BE85273

ALIGNMENTS

RESULT 1
LOCUS AL514158 977 bp mRNA EST 13-FEB-2001
DEFINITION AL514158 LTI_NFL006_P12 Homo sapiens CDNA clone CLOBB0052B08.5
ACCESSION AL514158
VERSION AL514158.1 GI:12777652
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 977)
AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1. 977
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBB0052B08"
/issue_type="Placenta"
/note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand CDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded CDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 321 a 188 c 213 g 253 t 2 others

Query Match 35.0%; Score 896; DB 105; Length 977;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 946; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	738	caaccctctatctacgaagatcatgtaccacacccggaagctctgcagatgaca	797
Db	1	CACCTCTGATGATTCAGGAAGTATGTTCACCAACCGGAGGCTGGCAGATGACA	60
Qy	798	gtttttatcacaggtgcgaagcgtgtgcattgcgaagctatgtcatgtgaagcagaag	857
Db	61	GTTTTATTCACAGCTGCAGCGCTGGCATTGGCAACCTATTGCATTGAAGCAGCAAG	120
Qy	858	gatgagcaaatatgttatgtctgcgaagaacccgacacacacccaactctggc	917
Db	121	GATGAGCAAAATATGTATGTATGTCTGCAGAACCGCCGACACATCCAAACTTGAAGC	180
Qy	918	acaatcctactctgtctgcgaagaatgcgaagcaattgcgaagaagccttgcattgt	977
Db	181	ACATCTATCTGCTGCTGCAGGAAGTATGAGCACTTGGAGAAAGGCTTGCATGTATT	240
Qy	978	gttgatgtgagagatgcagacagatcagctgcagctgcagtggaagacatcaagaattc	1037
Db	241	GTTGATGTGAGAGATGGAACAGCATCATGCTCAGTGGAAGACCATCAAGAAATTT	300
Qy	1038	ggaggaattgatctctgtgtaataatgcagctgcattgattgcacatatactggac	1097
Db	301	GGAGGAATGTGATGTGTAATAATGACAGTCCATTTGTTGACCAATCAATTTGAC	360
Qy	1098	acacctaccaaagatgtgatctgatgaagctgcagacacagagcacctacttgc	1157
Db	361	ACACCTACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	420
Qy	1158	tctaaagcatgtatctcattatgaaagagcaagttgtcatatccatcaatacagt	1217
Db	421	TCTAAACATGTATCTCTTATTTAAAGAGCAAACTGTCTCATATCTCAATATAGT	480
Qy	1218	ccaccactgaacccaatccagctgtgttcaacagacacgtgttataccattgcaag	1277
Db	481	CCACCACCTGAACCTTAATCCAGTTTGGTTCAACAGCATGTCTTATACATTTGTAAG	540
Qy	1278	tatgtatgtctatgtatgtgtctgtgaatgcagagaatlttaaggltgaattgcagtc	1337
Db	541	TATGTATGTCTATGTATGTCTGTGAATGACAGAAATTTAAAGGTGAATTTGCACTC	600
Qy	1338	aatgcattatgtgcttaaaagacacatacacactgtctatggtatgctgtggggagcct	1397
Db	601	AATGCATTATGTGCTTAAAGCCATACACCTCTCTATGATGTCTGGAGGAGCT	660
Qy	1398	ggtatcgaaagccagctgtagaagaagttgatatcatctgcagatcatcttc	1457
Db	661	GGTATCGAAAGCCAGGTGAGAAAGTTGATTCATTGACATGACGATATTTCCATTTTC	720
Qy	1458	caaaagccaaaagtttactgtgcaacttgcattgatgaaaatatcttaaaagaa	1517
Db	721	CAAAAGCCAAAAGTTTACTGCGCACTTGTATGTATGTAATTAATCTTAAAGAA	780
Qy	1518	ggaatagaagaatttgcagcttatagtcaatcaaccaggtcaatcttgcacagatttc	1577
Db	781	GGAATAGCAAAATTTTGAAGCTTATGCAATTAACACAGTCTCTTGGCAACGATTTTC	840
Qy	1578	ttcttaataaataccagaagcagcttgaagaagaattgtaataactgtgtgttcca	1637
Db	841	TTCTTAATGATGATACCCAGAGAGCTTATGCAAGRAGTGGATTAACCTGGCTGTCCA	900
Qy	1638	gaattcaagaagaagcaactgcagctgaacccaacacacagcttctg	1684
Db	901	GAATTTCAAGAGAGAACTGCACTGCAACCAAAACACAGTTCTGG	947

RESULT 2
LOCUS AL528484 861 bp mRNA EST 13-FEB-2001
DEFINITION AL528484 LTI_NFL003_NBC3 Homo sapiens CDNA clone CSDDC029YE05.5

```

prime, mRNA sequence.
ACCESSION AL528484.1 GI:12791977
VERSION AL528484.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS Li W.B., Gruber C., Jessee J. and Polayes D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Source
1..861
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSODC029YE05"
/clone_1b="LTI_NFL003_NBC3"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Peng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : filiang@life.com url :
http://fulllength.invitrogen.com"

BASE COUNT 270 a 172 c 188 g 226 t 5 others
ORIGIN
Query Match 31.3%; Score 802; DB 106; Length 861;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 802; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 atgttaaccaacacgggagctgtgcagatgtacagttttatcaacaggtcaacgct 821
DB 60 atgttaccacacacgggagctgtgcagatgtacagttttatcaacaggtcaacgct 119
QY 822 ggcattggcaaaagctatctcattgaaagcaaaagatgagagcaaatcttcttctgt 881
DB 120 ggcattggcaaaagctatctcattgaaagcaaaagatgagagcaaatcttcttctgt 179
QY 882 gcaaaagccgcccagacacatccaactctcagcacatcattctctgtaaga 941
DB 180 gcaaaagccgcccagacacatccaactctcagcacatcattctctgtaaga 239
QY 942 atggaagcgttgggaggaagcctgtgcattgtgtgagtgagagatgaacagcag 1001
DB 240 atggaagcgttgggaggaagcctgtgcattgtgtgagtgagagatgaacagcag 299
QY 1002 atcagtgctgcagtgagagaaagcattcaagaatttgaggaattgatattctgttaaat 1061
DB 300 atcagtgctgcagtgagagaaagcattcaagaatttgaggaattgatattctgttaaat 359
QY 1062 aatgcagtgctcattgattgacaaatattgacacacccctacaaagatgtgattc 1121
DB 360 aatgcagtgctcattgattgacaaatattgacacacccctacaaagatgtgattc 419
QY 1122 atgatagaagtgataacacccagagcactccttgatcctaaagatgtattccattat 1181
DB 420 atgatagaagtgataacacccagagcactccttgatcctaaagatgtattccattat 479
QY 1182 aaaaagagaaggttcctatatccctcaatcattcagtcacccctgaacctaatacagtt 1241

```

```

DB 480 AAAAAGACCAAGTGTCTCATATCTCAATATCAGTCCACCACTGAACCTTAATCCAGTT 539
QY 1242 tggttcaaaacagcactgctctatattaccattgtaattgataatgataatgattc 1301
DB 540 tggttcaaaacagcactgctctatattaccattgtaattgataatgataatgattc 599
QY 1302 ggaatgacgaagaatttaaaagttgaatttcagtcacacgtcttaagcctaaacagc 1361
DB 600 ggaatgacgaagaatttaaaagttgaatttcagtcacacgtcttaagcctaaacagc 659
QY 1362 atacaacactgctcattgataatgctgtggaagacccgtgatacgaagaagcgttaga 1421
DB 660 atacaacactgctcattgataatgctgtggaagacccgtgatacgaagaagcgttaga 719
QY 1422 gttagataatcagcagatgacacatattcatttccaaagacaaagatttactgac 1481
DB 720 gttagataatcagcagatgacacatattcatttccaaagacaaagatttactgac 779
QY 1482 aacttgcattgataatgataatcttaaaagaaagaaagaaatcttgacgttat 1541
DB 780 aacttgcattgataatgataatcttaaaagaaagaaagaaatcttgacgttat 839
QY 1542 gcaatgaacacaggtcattcctt 1563
DB 840 gcaatgaacacaggtcattcctt 861

RESULT 3
BE893726 602 bp mRNA EST 20-OCT-2000
LOCUS BE893726
DEFINITION 601366381 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3921920 5',
mRNA sequence.
ACCESSION BE893726
VERSION BE893726.1 GI:10355379
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1998)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LHM9755 row: k column: 09
High quality sequence stop: 602.

FEATURES
Source
1..602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3921920"
/clone_1b="NIH_MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site:1. NotI;
Site:2. SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 2 kb. Library constructed by Life
Technologies."

BASE COUNT 196 a 109 c 129 g 168 t
ORIGIN
Query Match 22.8%; Score 584; DB 141; Length 602;
Best Local Similarity 100.0%; Pred. No. 8.6e-298;

```

Matches 584; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1226 gaacctaataccaggttggttcaaacagcactgtgttataccattgctatgtagt 1285
 |||||||
 Db 19 GAACCTAAATCCAGTTGGTTCAAAAGCAGCTGTGCTTATACCATGCTTAAGTATGCTAT 78
 |||||||

QY 1286 gctatgtatgtcttggaatgagcaagaatttaagaagtgaatttcagtcattg 1345
 |||||||
 Db 79 GTTATGTATGTCTTGGAATGGCAGAGAAATTTAAAGGTGAATTCAGTCAATGCTATT 138
 |||||||

QY 1346 atgagctaaacagcacaacacactgctatgatatgctggaagacctgtatcga 1405
 |||||||
 Db 139 ATGCGCTAAACAGCCATACACACTGCTGTATGATATGCTGGAGAGACTGGTATCGA 198
 |||||||

QY 1406 aagccagtgtagaagaagtgtatatcatctgcagatgcagcatatccattccaagaagc 1465
 |||||||
 Db 199 AACCCAGTGTAGAAAGTTATATCATTTGCAATGCAATTCATTTCCATTTTCCAAAAAGCC 258
 |||||||

QY 1466 aaaaagtttactgagcaacttgatcatgtagaataatcttaaaagaagaagtaga 1525
 |||||||
 Db 259 AAAAAGTTTACTGGCACTTTGTGATGATAAATATTTTAAAGACAGAGAAATAGA 318
 |||||||

QY 1526 aaatttgacgttatagtcaatlaaacaggtcaltcccttgcaaccagattctctaga 1585
 |||||||
 Db 319 AAATTTTGAAGCTTTATGCAATTAACAGAGTCAATCTTTCACACAGATTTCTTCTAGA 378
 |||||||

QY 1586 tgaatacccaagagcagttgcaagaagaagtgtgaatacaactgtgtgcttccagaattcaa 1645
 |||||||
 Db 379 TGAATACCCGAAAGCAGTGTGCAAGAAAGTGAAATCAACGCTGCTTCAGAAATTCAA 438
 |||||||

QY 1646 agagaagaacactgcagctgcaacaaacagcttctgagcgtgtggaagaagaattag 1705
 |||||||
 Db 439 AGAAGAGAAACTGACAGCTGCACACCAAAACACAGCTTCTGGAGCTGTGAAAGAACATTTAG 498
 |||||||

QY 1706 aattgttaagagactctctcagtgatgatgtgtttaaagcacaacgaactatctatcgtt 1765
 |||||||
 Db 499 AATTTGTTAAGAGACTCTCTCAGTGATGTGTTGTTAAAGCAGCTCAACCATCATCATGTT 558
 |||||||

QY 1766 tgaactctccgtgtgaagatggtgtgacagtggttcttgcattcga 1809
 |||||||
 Db 559 TGAATCTCCGTTGAAGATGTTGGCAGCTGTTCTTGATCTGA 602
 |||||||

RESULT 4
 LOCUS BF699109 887 bp mRNA EST 22-DEC-2000
 DEFINITION 602126730F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4283671 5',
 mRNA sequence.
 ACCESSION BF699109
 VERSION BF699109.1 GI:11984517
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 887)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONETECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.llnl.gov
 Plate: L1CM115 row: 1 column: 08
 High quality sequence stop: 611.
 Location/Qualifiers
 1. 887

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4283671"
 /clone_1ib="NIH_MGC_56"
 /tissue_type="primitive neuroectoderm"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: brain; Vector: pMDR-LIB (Clontech); Site_1:
 SfiI (ggcgagctggcc); Site_2: SfiI (ggcgatagggc);
 Double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGGCATTATGGC-3' and 3' adaptor
 sequence: 5'-ATTCTAGAGCGCGAGCGCCGACATG-dT(30)BN-3'
 (where B = A, C, G, or G and N = A, C, G, or T). Average
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

BASE COUNT 251 a 185 c 227 g 224 t

ORIGIN

Query Match 19.5%; Score 500; DB 168; Length 887;
 Best Local Similarity 100.0%; Pred. No. 2.8e-253;
 Matches 500; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1114 tgaatcgtatgatagaagtgtaaacaccagagcacctacccttgatcctaaagcatgtattc 1173
 |||||||
 Db 12 TGATCTGTATGATGAAAGTGAACACAGAGGACCTTACCTTGATCTTAAAGATGTATTC 71
 |||||||

QY 1174 cttaattgaaaagaagaaggttgctcatalccctcaatatcagtcacacgaactaa 1233
 |||||||
 Db 72 CTTATTTGAAAAAGAGCAAGTTGCTCATATCTCTCAATATACAGTCCAGCTGAACCTAA 131
 |||||||

QY 1234 atccagtttgctcaaacacagcactgtgtctataccattgttaagtgtatgtctatgt 1293
 |||||||
 Db 132 ATCCAGTTTGCTTCAACACAGCAGCTGTCTATACCATGCTAAGTATGATGTATGTGT 191
 |||||||

QY 1294 atgtgcttgtaatgagcagaagaatttaagtgtaattgcagtcacatgcatatgacct 1353
 |||||||
 Db 192 ATGTGCTTGGAAATGGCGAAGAAATTTAAAGTGAATTTGCAAGTCAATTTATGGCCTA 251
 |||||||

QY 1354 aaacagccatacacacactgctgctatgatatgctcggaggaagcctgtgatatgaagccagt 1413
 |||||||
 Db 252 AAACAGCATACACACTGCTGCTATGATATGCTGGAGGACCTGATATGAAGCCAGT 311
 |||||||

QY 1414 gtagaaagtgtatcatcatcagatgcagatcagcatatccatttccaaaagccaaaagtt 1473
 |||||||
 Db 312 GTAGAAAGTTGATATCATTTGCAAGATGCACATATTCATTTCGCAAAAGCAAAAGTT 371
 |||||||

QY 1474 ttaactggcaacttgtcatctgatagaataatcttaaaagaagaagaataagaatttg 1533
 |||||||
 Db 372 TTACTGGCAACTTGTTCATGATGAAGAAATATCTTAAAGAAAGAAAGCAATAGAAATTTTG 431
 |||||||

QY 1534 agcttatagtcaatlaaacccaggtcatccttgcaaccagattcttcttgatgaataacc 1593
 |||||||
 Db 432 ACGTTATGCAATTAACCGAGTCATCTTTGCAACCAAGATTTCTTCTTGATGAATAC 491
 |||||||

QY 1594 cagaagcagttgacaagaaga 1613
 |||||||
 Db 492 CAGAGCAGTTTACCAAGAAA 511
 |||||||

RESULT 5
 LOCUS AW152065 620 bp mRNA EST 03-NOV-1999
 DEFINITION xf74a01.x1 NCL CGAP Gas4 Homo sapiens cDNA clone IMAGE:2623752 3',
 similar to WP:CI7610.8 CE16861 FAT-3; ALCOHOL DEHYDROGENASE ;, mRNA
 sequence.
 ACCESSION AW152065
 VERSION AW152065.1 GI:6199963
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 620)
 NCI-CCAP <http://www.ncbi.nlm.nih.gov/nciccap>.
 TITLE
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL
 Unpublished (1997)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CCAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
www-bio.lnl.gov/db/ftp/image/image.html
 Seq primer: -400p from Glbco
 High quality sequence stop: 416.
 Location/Qualifiers
 1..620
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone IMAGE:2623752
 /clone_lib="NCI-CCAP_Gas4"
 /tissue_type="poorly differentiated adenocarcinoma with
 signet ring cell features"
 /lab_host="DH10B"
 /note="Organ: stomach; Vector: pCMV-SPORT6; Site:1; Salt:
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.65 kb. Life Technologies catalog #:
 11549-011"
 BASE COUNT 162 a 144 c 101 g 209 t 4 others
 ORIGIN
 Query Match 19.2%; Score 491; DB 111; Length 620;
 Best Local Similarity 100.0%; Pred. NO. 1.0e-248;
 Matches 491; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1528 atttgagcattatgcaattaaacaggatcatcttgcacacagattctcttagatg 1587
 |||
 Db 494 attttgacgtttatcatatTAACAGGTCATCTTTCACACAGATTCTCTTAGATG 435
 |||
 Oy 1568 aatacccaagcagttagcaagaagtgaatcaactggtgctgtccagaattcaag 1647
 |||
 Db 434 aataccagaaagcagttagcaagaagtgaatcaactggtgctgtccagaattcaag 375
 |||
 Oy 1648 aagagaactgagctgcacacaaacacagcttctgagctgtggaagaacattagaa 1707
 |||
 Db 374 aagagaaactgcagctgcacacaaacacagcttctgagctgtggaagaacattagaa 315
 |||
 Oy 1708 ttgttaagaaacctcagtgatgtgtttaaagccactcaagaatctctgtttg 1767
 |||
 Db 314 ttgttaagaaacctcagtgatgtgtttaaagccactcaagaatctctgtttg 255
 |||
 Oy 1768 aacctccggtgaagaatggtgacggtgttcttgatctgaaagcaaggtgggaatg 1827
 |||
 Db 254 aacctccggtgaagaatggtgacggtgttcttgatctgaaagcaaggtgggaatg 195
 |||
 Oy 1828 tcggatataggaaagccttcagtcagcagatgtggtatgatactatgatact 1887
 |||
 Db 194 tcggatataggaaagccttcagtcagcagatgtggtatgatactatgatact 135
 |||
 Oy 1888 ttgttaaaatgttttcaggaagaaactaaacaaatgcatctgttcaaggaaattga 1947
 |||
 Db 134 ttgttaaaatgttttcaggaagaaactaaacaaatgcatctgttcaaggaaattga 75
 |||
 Oy 1948 agattaaagttacatgagccctagcaatcaatltgagaagtaatacagatgaatg 2007
 |||
 Db 74 agattaaagttacatgagccctagcaatcaatltgagaagtaatacagatgaatg 15

Oy 2008 ccagactgtga 2018
 |||
 Db 14 CCAGACTGTGA 4
 |||
 RESULT 6
 BE892079 963 bp mRNA EST 20-OCT-2000
 LOCUS 601434975F1 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:3919890 5',
 DEFINITION
 mRNA sequence.
 BE892079
 ACCESSION BE892079.1 GI:10352047
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 human.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 963)
 NIH-MGC <http://imgc.ncbi.nlm.nih.gov/>.
 TITLE
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL
 Unpublished (1999)
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: ATCC/DCTD/DP
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://imgc.lnl.gov>
 Plate: LLM9750 row: f column: 19
 High quality sequence stop: 624.
 Location/Qualifiers
 1..963
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone IMAGE:3919890
 /clone_lib="NIH-MGC_72"
 /tissue_type="melanotic melanoma"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: skin; Vector: pCMV-SPORT6; Site:1: NotI;
 Site:2: SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 2 kb. Library constructed by Life
 Technologies."
 BASE COUNT 302 a 205 c 218 g 238 t
 ORIGIN
 Query Match 19.1%; Score 489; DB 141; Length 963;
 Best Local Similarity 100.0%; Pred. NO. 1.9e-247;
 Matches 489; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1254 cacttgcttatacattgctaagatggtatgctatgcttgaatggcagaa 1313
 |||
 Db 1 CACTGCTTATACCATGCTCTAAGTATGATGTATGCTATGCTGTGATGACAGAA 60
 |||
 Oy 1314 gaatttaaggtgaattgcaatgcaatgcaatgcaatgcaatgcaatgcaatgcaatg 1373
 |||
 Db 61 GAATTTAAAGGTGAATTGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 120
 |||
 Oy 1374 gctatgatactgctggggagacctggtatcgaaacccagtgagaagaattgatact 1433
 |||
 Db 121 GCTATGAGATATCTCGGAGGAGCACTGTATCGAAACCAAGTGAAGATTGATATCAAT 180
 |||
 Oy 1434 gcagatgcagcatatctcatttccaaagccaaagcttcttaccggaactgtcat 1493
 |||
 Db 181 GCAGATGCAGCATATTTCCATTTTCCAAAGCCAAAGATTCTTACTGCGCACTTTGTCAAT 240
 |||
 Oy 1494 gatgaataatctttaaagaagaagaatgagaataattgacgtttatgcaatlaaaca 1553
 |||
 Db 241 GATGAATAATCTTTAAAGAGAGAGAGATAGAAATTTTGACGTTTATGCAATTAAACA 300

Accession	Sequence	Position
Oy 1554	ggtatccctcttgcaaacaggtcttcctcttagatgaaataccagaagaacatttagcaagaa	1613
Db 301	ggctatccctttggcaccacagattttcttctttagatgaattcccgaaacagtttagcaagaa	360
Oy 1614	gtggaatcaaacctggtgctgtgtccagaatltcaagaagaagaactgcaagctgtgcaacaa	1673
Db 361	gtggaatcaaacctggtgctgtgtccagaatltcaagaagaagaactgcaagctgtgcaacaa	420
Oy 1674	ccacggtctctggaagctgttgaagaagaacattagaattgtttaagaaactcttcagtgatgat	1733
Db 421	ccacggtctctggaagctgttgaagaagaacattagaattgtttaagaaactcttcagtgatgat	480
Oy 1734	gttggttaaa 1742	
Db 481	gttggttaaa 489	

RESULT	7
LOCUS	AM966207
DEFINITION	AM966207 541 bp mRNA
VERSION	ESTJ78280 MAGE resequences, MAGI Homo sapiens CDNA, mRNA sequence.
KEYWORDS	AM966207.1 GI:8156043
SOURCE	EST.
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 541)
	Hegde, P., Qi, R., Abernathy, K., Dharap, S., Gaspar, R., Gay, C., Holm, I.E., Saeed, A.I., Sharov, V., Lee, N.H., Yeatman, T.J. and Quackenbush, J.
TITLE	Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element CDNA microarray
JOURNAL	Unpublished (2000)
COMMENT	Contact: John Quackenbush The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 3528 Fax: 301 838 0208 Email: john@tigr.org Plate: 224
	Seq primer: Reverse.
FEATURES	Location/Qualifiers
source	1..541
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone_lib="MAGE resequences, MAGI"
	/note="Vector: pBluescriptSM"
BASE COUNT	172 a 110 c 122 g 137 t
ORIGIN	

Query Match	17.3%	Score 442	DB 122	Length 541
Best Local Similarity	100.0%	Prod. NO. 1.5e-222		
Matches 442	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	745	tcgatctcgaagaagtcattgtaaccaacaccggyagctgycagaatgtacagtttta	804	
Db	1	TCGTATCTACCAAAAGTCATGTACCACACCCGGAGCTGCAGAGATGTACAGTTTTA	60	
QY	805	tcacaggtgcgaagccgttcgacttggcgaagcattgcacttgaagaacgagaagttgag	864	
Db	61	TCACAGGTGCAGCCGTGTGCATTTGCCAAGCTATTGCATTTGAACACAGCAAAAGATGGAG	120	
QY	865	caaatattgtatttgcgtcgaagaccgcgcacacatccaaaactcttaggcacaatct	924	
Db	121	CAAAATATGTATTGCTCTCCAAAAGCCGCCACCATCCCAAAACTTTAGCCACAACT	180	
QY	925	atactgtcgtcgcgaagaattgaagcagtttgtaggaaaagcccttcgcagtatttctgatg	984	
Db	181	ATACGCTGCTGAAGAAATTTGAAGCTATTGGAGGAAAGCCCTTGCCATCATTTGTTGAG	240	

OY	985	tggagatgtaacgcgcgatatacgtcgccagcttgagaagaccataagaatttgaaggaa	10444
Db	241	TGAAGATGTAACGCAGCATATGTCTCGAGTGGAGAAACCATTCAAGAAATTTGGAGGAA	300
OY	1045	ttgatattcgtgtaaataatgtaacgtgccattagtttgaccaatacatatggacacctta	1104
Db	301	TTGATATTTCGTGAATAAATAGCCAGTGCCATTAGTTTGACCAATTCATTGGACACACTTA	360
OY	1105	ccaagagatttgatctgatagtatgaacytgacaacccaagagcaacttaacttgatctaag	11644
Db	361	CCAAGACATTGGATCTGCATGATGATGMAAGGTGAACACCGAGAGGCCACTTATGATCTTAAG	420
OY	1165	catgatttccttatttgaaaaa	1186
Db	421	CATGATTCCTTATTTGAAAAA	442

RESULT	8
AA315514	
LOCUS	
DEFINITION	AA315514 487 bp mRNA EST 19-APR-1997
ACCESSION	ESR187305 COLON CARCINOMA (HCC) CELL LINE II Homo sapiens cDNA 5
VERSION	AA315514
KEYWORDS	end, mRNA sequence.
SOURCE	AA315514.1 GI:1967863
ORGANISM	EST.
REFERENCE	human.
AUTHORS	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
	1 (bases 1 to 487)
	Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult

TITLE	JOURNAL	MEDLINE	COMMENT
Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence	Nature 377 (6547 suppl), 3-174 (1995)	96026280	Other-ESTs: THC130362 Contact: Kerlavage, AR Bioinformatics The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850 USA Tel: 3018699056 Fax: 3018699423 Email: akervavet@tigr.org For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/ghi/ghi.html) Seq primer: M13 Reverse.
Location/Qualifiers			
1. 487			
/organism="Homo sapiens"			
/db_xref="ATCC (inhost):110493"			
/db_xref="taxon:9606"			
/clone_lib="Colon carcinoma (HCC) cell line II"			
/tissue_type="colon"			
/cell_type="KM12C"			
/cell_line="KM12C(HCC)-parental human colon carcinoma ;Dukes B2"			

```

/note="Organ: colon; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"
BASE COUNT      141 a      110 c      115 g      119 t      2 others
ORIGIN
Query Match      16.4%; Score 419; DB 5; Length 487;
Best Local Similarity 99.8%; Pred. No. 2.2e-210;
Matches 469; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 714 gctgcggccggcgctgtgcgcgcacccctctctatctacgaagtcagtaccacac 773
Db 18 gctgcggccggcgctgtgcgcgcacccctctctatctacgaagtcagtaccacac 77
Oy 774 accggaagcttgcgcagatgtacagttttatcacagtgcaagccgttgcata 833
Db 78 accggaagcttgcgcagatgtacagttttatcacagtgcaagccgttgcata 137
Oy 834 gctatctcattgaaagcagcaagagatggagcaaatatgttattctgcgaagccgc 893
Db 138 gctattgcatgaaagcagcaagagatggagcaaatatgttattctgcgaagccgc 197
Oy 894 cagccacatcccaaatcttgaagcaaatctatctctgcctgaaagaaatgaaagcagt 953
Db 198 cagccacatcccaaatcttgaagcaaatctatctctgcctgaaagaaatgaaagcagt 257
Oy 954 ggaaggaagccttgcatactgtatgttgaatgagatgaaagcagacagtcagtcga 1013
Db 258 ggaggaagccttgcatactgtatgttgaatgagatgaaagcagacagtcagtcga 317
Oy 1014 gtgaggaagccttgaagaaatgttgagaaatgttatctgttaataatccagtcgc 1073
Db 318 gtgaggaagccttgaagaaatgttgagaaatgttatctgttaataatccagtcgc 377
Oy 1074 attaatcttgaacacatcttgcacacacccacaaagaaatgaaatgaaatgaaatg 1133
Db 378 attaatcttgaacacatcttgcacacacccacaaagaaatgaaatgaaatgaaatg 437
Oy 1134 aacaccagagcagcctacctcgtcatctaaagacatgattccttattgaa 1183
Db 438 AACACCAGAGGACCTACCTGATCTAAGCATGATGATTCCTTATTGAA 487
RESULT 9
AM299683 692 bp mRNA EST 18-JAN-2000
LOCUS x942607.x1 NCI_CGAP_KiD1 Homo sapiens CDNA clone IMAGE:2772314 3'
DEFINITION similar to FR:019066 019066 OVARIAN STEROL CARRIER PROTEIN 2 ;
ACCESSION AM299683
VERSION AM299683.1 GI:6709360
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 692)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40up from Gibco

```

```

High quality sequence stop: 466.
Location/Qualifiers
1..692
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="2772314"
/clone_lib="NCI_CGAP_KiD1"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_KiD3 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 1322376-1323911, 1456007-1456775, and
1500552-1502855). Subtraction by Bento Soares and M.
Fatima Bonaldo."
BASE COUNT      227 a      130 c      93 g      242 t
ORIGIN
Query Match      16.3%; Score 418; DB 113; Length 692;
Best Local Similarity 100.0%; Pred. No. 7.6e-210;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 2045 aggaataataaaaaaagtcagctcctacatgctcaaaaaagtaaaaaagctcaacagt 2104
Db 476 AGGAAATATTAATAAAAAAGTCAGCTCTATCTCAAAAAAGTAAAGCAACAGTT 417
Oy 2105 aaaaataatgtttcttcttctcctgtatataatgaagatgacgtttcttctgaa 2164
Db 416 AAAAACTAATGTGTTGTTCTTCTTCTGTTATATTAAGATATGACGTTGTTCTGCA 357
Oy 2165 aagaatagaattgtctctaaagaacttgaacttgaacttgaacttgaacttgaactt 2224
Db 356 AAGAATAGAAATTTGCTCTAAAGAACTGAATTAATTAATTAATTAATTAATTAATTA 297
Oy 2225 caaagctcattgaagtggaattcaagacagctcgtgttttataatcaagagtttaa 2284
Db 296 CATAGCTTCAATTAAGTGAATCTTAAGACAGCTGCTGTGTTTAATTAATTAATTAAT 237
Oy 2285 ccccttgagccttaccatctcactcgtctcttcccaagaagtaatttggggagaca 2344
Db 236 CCCCTTGAGCCTTACATCTCATTCACCTGCTTCTCCAGAAAGTATTTTGGGGAGACA 177
Oy 2345 gtcagatcaagcaatgaatagctcttcaaatcttctgtcatgtaaatgaaagctag 2404
Db 176 GTCAGATCAAGCAGTAATTAAGCTCTTCAAAATCTCTTGCATGTAATTAATTAATTA 117
Oy 2405 tctgttttaaaattttagtttctgtgattgatactaaagaatccttaagtattt 2462
Db 116 TCTGTTTAAATTTTAAATTTTGAATTTGATGTATATGTAATTAATTAATTAATTAAT 59
RESULT 10
BF979524 829 bp mRNA EST 23-JAN-2001
LOCUS 602288151F1 NIH_MGC_97 Homo sapiens CDNA clone IMAGE:4373773 5'
DEFINITION mRNA sequence.
ACCESSION BF979524
VERSION BF979524.1 GI:12346843
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 829)
AUTHORS NIH-MGC http://mgc.ncbi.nlm.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550

```


Query Match	12.9%	Score 330;	DB 157;	Length 484;
Best Local Similarity	100.0%	Pred. No. 3.2e-163;		
Matches 330;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```

/organism="Homo sapiens"
/db_xref="CDB:413643"
/db_xref="EAXON:9606"
/clone="IMAGE:41102"
/clone_lib="Scars infant brain INTB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10 (ampicillin resistant)"
/notes="Organ: whole brain; Vector: Lambda BA; Site_1: Not
I; Site_2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5'
AACGGAGAGATTTCGCCGCCGACGAGATTTTTTTTTTTT 3']";
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned

```


Email: genexpress@genethon.fr
 strand(+), single read
 Warning!! There has been some controversy in the scientific
 literature over
 possible contamination of this library by yeast and prokaryotic
 sequences.
 See Science 259:1677-1678 (1993).

FEATURES
 source
 Location/Qualifiers
 1..340

/organism="Homo sapiens"
 /db_xref="GDB:D05690E"
 /db_xref="taxon:9606"
 /clone="28F11"
 /clone_lib="STRATAGENE Human skeletal muscle cDNA library,
 cat. #936215."
 /sex="female"
 /tissue_type="skeletal muscle"
 /dev_stage="19 years"
 /note="Organ: leg muscle; Vector: Lambda ZAPII; Tissue
 from female, 19 years old, normal leg muscle. Cloning
 vector is Lambda ZAPII, in vivo excision from lambda ZAPII
 to pluscript SK(+). Genexpress library reference is B. "
 BASE COUNT 106 a 47 c 59 g 127 t 1 others
 ORIGIN

Query Match 12.4%; Score 317; DB 159; Length 340;
 Best Local Similarity 100.0%; Pred. No. 2.5e-156;

Matches 317; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2128 cctgtatattttaaaggaatgcacgcttcttctgaaagaatagatttgcctctaaa.2187
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 CCTGTATATTATTAAGATATGCACGTTGTTCTGAAAGAATGATTTGTCCTTAAAA 60
 QY 2188 gaactgaattgtaattaaatggcaagctaaatcaacaataagcttcattaaatggatt 2247
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 GACTTGAATTTGTAATTAATGCGCAAGCTATTCACAACATAGCTTCATTAGTGGATT 120
 QY 2248 ctaagaagtcctgcttcttataattcaagggttaaccccttgagccttaacatcatt 2307
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 121 CTAAAGACACTGCTGTTTATTTCAAGGGTTTAAACCTTTGAGCCTTACATCTCATTT 180
 QY 2308 cactgctctcccaagaagaatttttggggagacagtcagatcaagcagtaaaattag 2367
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 181 CACTGCTTTCTCCAGAAAGATTTTGGCGGACAGTCAGATCAAGCAGTAAAAATTAG 240
 QY 2368 ccttcaaatctcttgcacatgtaaaatgaagcagtcagtcgtttaaattttaattt 2427
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 CTCTTCAATCTCTCTGTCATGTAAATGAAGCTAGTCTGTTTAAATTTTACTTTT 300
 QY 2428 ggaactgatataactaata 2444
 ||||||||||||||||
 Db 301 GGATTGTATATCAATGA 317

Search completed: June 12, 2001, 13:10:48
 Job time: 5425 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 10:40:42 : Search time 169.81 Seconds
(without alignments)
8804.334 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561
Sequence: 1 aggcagagatgcaaaagca.....gsgmggraswvawvrammc 2561

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: N_Geneseq_0401.*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	556.2	21.7	4951	19	V27206 Vector pLtrAP3 DNA
2	400	15.6	633	18	T97159 SR alpha enhancer/
3	354.2	13.8	633	19	V35786 Sequence of the sp
4	259.4	10.1	755	10	N92604 Recombinant DNA ve
5	259.4	10.1	9045	17	T42902 DNA sequence which
6	259.4	10.1	9047	5	N40080 Sequence complemen
7	249.4	9.7	720	10	N90825 Human T lymphotrop
8	249.4	9.7	720	15	O68398 HTLV-1 LTR genomic
9	249.4	9.7	720	18	T47852 HTLV-1 long termin
10	230	9.0	1675	14	O50836 HTLV-1 p21x cDNA.
11	230	9.0	1866	14	O50835 HTLV-1 tax/rex spl

12	181.4	7.1	422	13	O25258 HindIII fragment O
13	180.4	7.0	198	21	C08959 Human secreted pro
14	167.6	6.5	7287	19	V02042 Plasmid pWRG3169 e
15	166.2	6.5	5068	20	X84027 MMP9 promoter GFP
16	166	6.5	3796	21	A27831 Vector plasmid PCM
17	166	6.5	6253	20	X08454 AAV vector sequenc
18	166	6.5	6280	20	X08455 AAV vector sequenc
19	166	6.5	6280	20	X08456 AAV vector sequenc
20	166	6.5	6295	19	V02043 Plasmid pWRG3196 e
21	166	6.5	6981	21	X45931 Nucleotide sequenc
22	166	6.5	7380	20	X84028 MMP9 promoter beta
23	166	6.5	7612	21	Z39629 DNA sequence of pL
24	166	6.5	7852	17	T27556 Shuttle vector pad
25	166	6.5	7892	13	O30906 PSAB132. Syntheti
26	166	6.5	7897	17	T27555 Shuttle vector pad
27	166	6.5	8236	18	T47202 Recombinant adenov
28	166	6.5	8509	18	T59271 Plasmid pAV.CMVlac
29	166	6.5	8509	20	X33862 AV.CMVlac2 cis pla
30	166	6.5	9077	19	V09005 Vector containing
31	166	6.5	9077	19	V09006 Vector containing
32	166	6.5	9077	21	Z45251 Expression vector
33	166	6.5	9077	21	Z45253 Expression vector
34	166	6.5	10398	17	T15286 Ad.AV.CMVlac2 hybr
35	166	6.5	10398	19	V22130 cDNA sequence of p
36	166	6.5	10398	19	V22139 cDNA sequence of p
37	166	6.5	10398	19	V22139 cDNA sequence of p
38	166	6.5	12135	18	T60555 Plasmid pADCMVgag-
39	166	6.5	19307	17	T27558 Shuttle vector pad
40	166	6.5	36538	18	T60558 Shuttle vector trans-
41	164.6	6.4	304	20	V86402 EST clone AR253.
42	163.6	6.4	7286	20	X57224 WO 9923223 Seq ID
43	163.6	6.4	7938	20	X57235 WO 9923223 Seq ID
44	163	6.4	5141	16	T51451 Plasmid pSV17.ID.L
45	163	6.4	5141	16	O94114 Expression vector

ALIGNMENTS

RESULT	1
ID	V27206 standard; cDNA; 4951 BP.
XX	XX
AC	V27206;
XX	XX
DT	12-OCT-1998 (first entry)
XX	XX
DE	Vector pLtrAP3 DNA sequence.
XX	XX
KW	Secreted protein; membrane-associated protein; protein secretion;
XX	XX
OS	signal peptide; alkaline phosphatase; pLtrAP3; vector; reporter; ss.
XX	XX
OS	Homo sapiens.
XX	XX
FH	Key
FT	mat_peptide
FT	Location/Qualifiers
FT	1313..2782
FT	/*tag= a
FT	/product= human placental alkaline phosphatase
XX	XX
PD	28-MAY-1998.
XX	XX
XX	MO9822491-A1.
XX	XX
XX	06-NOV-1997; 97WO-US20201.
XX	XX
XX	19-NOV-1996; 96US-0752307.
XX	XX
PA	(MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
XX	XX
PI	Gearing DP, Levinson DA, McCarthy SA;
XX	XX
DR	WPI: 1998-312407/27.
DR	P-PSDB: W55047.

XX Identifying DNA encoding mammalian protein having signal sequence.
 PT by cloning in reporter gene vector without signal sequence.
 PT transforming bacteria and mammalian cells, and identifying cells
 PT that express reporter
 XX
 PS Claim 6; Fig 2; 50pp; English.
 XX
 CC This is the DNA sequence of vector pLTPA3. This mammalian
 CC expression vector contains a cDNA encoding human placental alkaline
 CC phosphatase (AP, see W55047) lacking a signal sequence. When
 CC pLTPA3 is transfected into a mammalian cell line, such as COS7
 CC cells, AP protein is neither expressed nor secreted since the AP
 CC cDNA of pLTPA3 does not encode a signal peptide or a membrane
 CC anchor sequence. However, insertion of a cDNA encoding a signal
 CC peptide sequence into pLTPA3 facilitates the expression and
 CC secretion of AP upon transfection of the DNA into mammalian cells.
 CC The presence of AP activity in the supernatants of transfected
 CC cells therefore indicates the presence of a signal sequence in the
 CC cDNA of interest. This forms the basis of a novel method for the
 CC identification of genes encoding novel proteins having a signal
 CC sequence, i.e. secreted or membrane-associated proteins of
 CC potential therapeutic value. An isolated clone (see V27707),
 CC designated eth0018f2, was isolated using the novel method and
 CC encodes a human neural adhesion protein (see W55045) protein having
 CC multiple, consecutive IgG domains. The method is very sensitive
 CC and is suitable for high throughput screening techniques and
 CC automation.
 CC
 XX
 SO Sequence 4951 BP; 1151 A; 1410 C; 1303 G; 1087 T; 0 other;

Query Match 21.7%; Score 556.2; DB 19; Length 4951;
 Best Local Similarity 96.2%; Pred. No. 1.8e-78;
 Matches 656; Conservative 0; Mismatches 14; Indels 12; Gaps 8;

QY 1 aggcagagaagatgcaagatgcatctcctaattagtcagaacaacatgcccgcctc 60
 DB 131 aggcagagaagatgcaagatgcatctc-aattagtcagc-aaccatagtcctc-gccctc 187
 QY 61 aactccgcatccgccccttaactccgncaggtccgcccattccgcccctatgac 120
 DB 188 aactccgcatccgccccttaactccgccc--agttccgcccattccgcccctatgac 245
 QY 121 tgactaattttttattatgacaga-gccgagagccgctcgtgctcgtatcag 179
 DB 246 tgactaattttttattatgacagagccgagccgctcgtgctcgtatcag 305
 QY 180 aagtagtagagagagcttttttgagagcctaggtctttgcaaaaagctcct-cgatcgag 238
 DB 306 aagtagtagagagagcttttttgagagcctaggtctttgcaaaaagctcctcgatcgag 365
 QY 239 ggttcgcatctcctcctcaagcgccgcccctactactgagcgccatccagccggt 298
 DB 366 ggttcgcatctcctcctcaagcgccgcccctactactgagcgccatccagccggt 425
 QY 299 gagtcggtctcgcgctcccgctcgcgtgctgctcctcgaactcgctcgcgtcaggt 358
 DB 426 gagtcggtctcgcgctcccgctcgcgtgctgctcctcgaactcgctcgcgtcaggt 485
 QY 359 aagtttaaaagctaggtcgagagccggtttgctggcgctcctcttgagagctactag 418
 DB 486 aagtttaaaagctaggtcgagagccggtttgctggcgctcctcttgagagctactag 545
 QY 419 acccagcggcgtctcagagcttgctgacccgtgctgacccactacgtactgtttc 478
 DB 546 acccagcggcgtctcagagcttgctgacccgtgctgacccactacgtactgtttc 605
 QY 479 agtttcctgcttgcgcgcttacagatcccaagctctgaaaaacagagaagttaactgta 538
 DB 606 -gtttctctgtctgcgcgcttacagatcccaagctctgaaaaacagagaagttaactgta 664
 QY 539 agtttagctcttctgtctcttattcaggtcccgatccgg-----cgtggtgacaaatca 594

DB 665 agtttagctcttctgtctcttattcaggtcccgagcccgccgagctcgatccaaatct 724
 QY 595 aagacagctcctcctcagagagatgtgctccttactctcagagctgtaagagagtgtaactc 654
 DB 725 aagacagctcctcctcagagagatgtgctccttactctcagagctgtaagagagtgtaactc 784
 QY 655 tgctctaaagctcgaggaatc 676
 DB 785 tgctctaaagctcgaggaatc 806

RESULT 2

T97159

ID T97159 standard; DNA: 633 BP.

AC T97159;

DT 11-MAY-1998 (first entry)

DE SR alpha enhancer/promoter.

KW Vaccine; B-cell malignancy; lymphoma; leukaemia; tumour;

KW gene amplification; immunotherapy; therapy; SV40; promoter;

KW enhancer; ds.

OS Chimeric - Rhesus macaque polyoma virus.

OS Chimeric - Human T cell Leukaemia virus type 1.

XX WO9741244-A1.

XX 06-NOV-1997.

XX 25-APR-1997; 97WO-US07039.

XX 06-DEC-1996; 96US-0761277.

XX 01-MAY-1996; 96US-0644664.

XX (GENI-) GENITOPE CORP.

XX Denney DW;

XX WPI; 1997-549743/50.

XX Multivalent vaccine to treat B cell lymphoma or leukaemia -

XX comprises at least 2 different recombinant variable regions of

XX immunoglobulin molecules derived from B cell lymphoma cells

XX Example 1; Page 104; 177pp; English.

XX This DNA sequence comprises the HindIII/XhoI fragment of the

XX SR alpha enhancer/promoter in plasmid pCDL-SR alpha 296. The

XX virus 1.5' untranslated sequences and the SV40 enhancer. It is

XX reported to increase expression from the SV40 enhancer/promoter by

XX 10-fold in host cells, and is active in a broad range of cell

XX types. The SR alpha enhancer/promoter has been utilised in

XX expression vectors designed for efficient expression of genes in

XX eukaryotic cells. The invention provides vectors and improved

XX methods for the expression and co-amplification of genes encoding

XX recombinant proteins in cultured cells. The methods permit the

XX isolation of cell lines which have co-amplified input recombinant

XX sequences which encode an amplifiable marker, one or more

XX expression vectors encoding a protein of interest and optionally a

XX selectable marker. The amplified cells provide large quantities of

XX recombinant proteins suitable for immunotherapy for treatment of

XX lymphomas and leukaemias. The methods permit the production of

XX custom vaccines, including multivalent vaccines that reflect the

XX degree of somatic variation found in a patient's tumour.

XX Sequence 633 BP; 114 A; 210 C; 155 G; 154 T; 0 other;

Key	Location/Qualifiers
FT	1..353
FT	/tag= a
FT	/label= U5
FT	354..381
FT	LTR
FT	/tag= b
FT	/label= R
FT	582..755
FT	/tag= c
FT	/label= U3
XX	
PN	JP01277489-A.
XX	
PD	07-NOV-1989.
XX	
PE	28-APR-1988; 88JP-0105677.
XX	
PR	28-APR-1988; 88JP-0105677.
XX	
PA	(GANK) GAN KENKYUKAI ZH.
XX	
DR	WPI; 1989-368596/50.
XX	
PT	Recombinant DNA vector - contains HTLV-I long terminal repeat ligated to
PT	part of U5, for increased chloramphenicol acetyltransferase activity.
XX	
XX	Disclosure; ; Japanese.
XX	
CC	HTLV-I LTR is inserted into pSV-CAT, the U3 region is lost, and the R-U5
CC	region gives 10-100 fold CAT expression. Vector can be integrated
CC	in a wide range of animal cells for high levels of expression.
XX	
SO	Sequence 755 BP; 161 A; 255 C; 178 G; 161 T; 0 other;
Query Match 10.1%; Score 259.4; DB 10; Length 755;	
Best Local Similarity 99.3%; Pred. No. 2.7e-32;	
Matches 271; Conservative 0; Mismatches 1; Indels 1; Gaps 1;	
QY	227 ggggctgcacatctctccttcacagcgccgcgcctactcagagccgacatccagccgg 296
DB	352 ggggctgcacatctctccttcacagcgccgcgcctactcagagccgacatccagccgg 411
QY	297 ttgaatcgatctctcgcgcctccgcgcctctgtggtgctctgaactgcgtccgctctag 356
DB	412 ttgaatcgatctctcgcgcctccgcgcctctgtggtgctctgaactgcgtccgctctag 471
QY	357 gtaagttaaaagctcaggtcgaagacgggcttctgcggcgctcccttgagagctact 416
DB	472 gtaagttaaaagctcaggtcgaagacgggcttctgcggcgctcccttgagagctact 531
QY	417 agaccacagcggtcttcacagcttgccttgaccccgctgttcacacttaagctcttgt 476
DB	532 agaccacagcggtcttcacagcttgccttgaccccgctgttcacacttaagctcttgt 591
QY	477 tcaagttctctctcgcgcgttaacagatccaa 509
DB	592 tc-gttctctctcgcgcgttaacagatccaa 623
RESULT 5	
T42902	
ID T42902 standard; DNA; 9045 BP.	
XX T42902;	
XX	
DT 16-JUN-1997 (first entry)	
XX	
DE DNA sequence which regulates expression in HTLV and HIV.	
XX	
KW gene expression; regulation; plasmid; viral infection;	
KW human T-cell leukemia; HIV; antiviral agent; detection; cancer;	
KW gene therapy; trsp; ss.	

XX	Homo sapiens.
OS	
XX	
FH	Key
FT	misc_feature
FT	Location/Qualifiers
FT	1..757
FT	/tag= A
FT	/label= LTR
FT	polyA_signal
FT	8584..8589
FT	/*tag= b
FT	misc_feature
FT	8278..9032
FT	/*tag= c
XX	
PN	MO9630522-A1.
XX	
PD	03-OCT-1996.
XX	
PP	19-MAR-1996; 96WO-JP00719.
XX	
PR	27-APR-1995; 95JP-0104299.
PR	24-MAR-1995; 95JP-0065559.
XX	
PA	(SHIO) SHIONOGI & CO LTD.
XX	
PI	Igarashi H, Okumura K, Orita S, Saiga A, Sakaguchi G;
DR	WPI: 1996-455367/45.
XX	
PT	DNA molecule with gene expression regulation activity - for use in
PT	e.g. treatment of human T-cell leukaemia and HIV, as antiviral agent
PT	and for detecting cancer
XX	
PS	Claim 2; Page 38-43; 77pp; Japanese.
XX	
CC	This sequence represents a DNA molecule with gene expression regulation
CC	activity. This sequence is used in a plasmid for regulation of gene
CC	expression, and treatment of viral infection pref. human T-cell leukemia
CC	and HIV. The plasmid also encodes a protein which is used as an antiviral
CC	agent, and also in a method for detecting cancer. The DNA molecule and
CC	protein have potential use in gene therapy, and the plasmid may also
CC	have potential use in the treatment of TSP.
XX	
SQ	Sequence 9045 BP: 2086 A; 3162 C; 1713 G; 2084 T; 0 other;
	Query Match 10.1%; Score 259.4; DB: 17; Length 9045;
	Best Local Similarity 99.3%; Pred. No. 3.5e-32;
	Matches 271; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
OY	237 ggggctgcacatcctcccttcacagcgccgccgacctagtaggcgccacacacgcgg 296
DB	
OY	352 ggggctgcacatcctcctcttcaacgcccgcgcgcgccttaactgaagcgcatcacgcgg 411
OY	297 ttgagtcggtctctgcgcgcctccgcgctgtgtgctctcctaactggaactgogctccgcgttag 356
DB	
OY	412 ttgagtcggtctctgcgcgcctccgcgctgtgtgctctcctaactggaactgogctccgcgttag 471
OY	357 gtaagtttaaagaagccaagttcgaagaccgggaccttgttcgcggcgcccccttgtagaacctact 416
DB	
OY	472 gtaagtttaaagaagccaagttcgaagaccgggaccttgttcgcggcgcccccttgtagaacctact 531
OY	417 agaatcagcgcgctctccacagcttttgcttgacacctgtctgttcataactatactatttgtt 476
DB	
OY	532 agaatcagcgcgctctccacagcttttgcttgacacctgtctgttcataactatactatttgtt 591
OY	477 tcaagttctgtctctgcgcggttacagatcaa 509
DB	
OY	592 tc-gttctctgtctctgcgcggttacagatcaa 623
RESULT	6
N40080	
ID	N40080 standard; cDNA; 9047 BP.
XX	

[illegible][illegible]

CC	mRNAs encoding secreted proteins. No ORF has yet been conclusively			
CC	identified within the present sequence. The 5' ESTs were prepared from			
CC	total human RNAs or polyA ⁺ RNAs derived from 30 different tissues. EST			
CC	sequences usually correspond mainly to the 3' untranslated region (UTR)			
CC	of the mRNA because they are often obtained from oligo-dT primed cDNA			
CC	libraries. Such ESTs are not well suited for isolating cDNA sequences			
CC	derived from the 5' ends of mRNAs and even in those cases where longer			
CC	cDNA sequences have been obtained, the full 5' UTR is rarely included.			
CC	5' ESTs are derived from mRNAs with intact 5' ends and can therefore be			
CC	used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used			
CC	in diagnostic, forensic, gene therapy and chromosome mapping procedures.			
CC	They are used to obtain upstream regulatory sequences and to design			
CC	expression and secretion vectors.			
XX				
SQ	Sequence 198 BP; 47 A; 54 C; 51 G; 46 T; 0 other;			
Query Match 7.0%; Score 180.4; DB 21; Length 198;				
Best Local Similarity 99.5%; Pred. No. 4.9e-20;				
Matches 181; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
OY	711 tccgctgcgcgcgcgcgcgtctgcgcgcacccctccctgatactcaagaatgctacc 770			
Db	17 tctgctgcgcgcgcgcgcgtctgcgcgcacccctccctgatactcaagaatgctacc 76			
OY	771 aacaccggagagctgcgcagatgtacaagttttatcaagtgcaagccgtgcatggc 830			
Db	77 aacaccggagagctgcgcagatgtacaagttttatcaagtgcaagccgtgcatggc 136			
OY	831 aagatattgctttaaagcagcaagaagatgagaagaatatgtattgtgtcaagacc 890			
Db	137 aagatattgctttaaagcagcaagaagatgagaagaatatgtattgtgtgtcaagacc 196			
OY	891 gc 892			
Db	197 gc 198			
RESULT 14				
ID	V02042			
AC	V02042 standard; DNA: 7287 BP.			
XX	V02042:			
DT	08-JUN-1998 (first entry)			
XX				
DE	Plasmid pWRG3169 encoding murine interleukin-12.			
KW	Interleukin-12; IL-12; cytokine; growth factor; mouse;			
KW	plasmid pWRG3169; cancer; tumour; metastasis; gene therapy; ds;			
KW	cyclic; circular.			
XX				
OS	Chimeric - Mus musculus.			
OS	Chimeric - Cytomegalovirus.			
OS	Chimeric - Bos taurus.			
XX	Chimeric - Rhesus macaque polyoma virus.			
PH				
FT	Key			
FT	promoter			
FT				
FT	location/Qualifiers			
FT	1..628			
FT	/*tag= a			
FT	/note= "CMV promoter"			
FT	1DNA			
FT	629..810			
FT	/*tag= b			
FT	953..1673			
FT	CDS			
FT				
FT	/*tag= c			
FT	/product= p35 subunit			
FT	/note= "contains an intron"			
FT	1259..1331			
FT	/*tag= d			
FT	1797..2024			
FT	polyA_site			
FT				
FT	/*tag= e			
FT	/note= "bovine growth hormone polyA site"			
FT	2110..2737			
FT	promoter			

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run On: June 12, 2001, 09:24:42 ; Search time 1765.19 Seconds
(Without alignments)
12674.640 Million cell updates/sec

Title: US-09-464-039-8
Perfect score: 2561
Sequence: 1 aggcagaagatgcagaagca.....gsgmgtasmmwawrrammc 2561

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: gb_est41:*
42: gb_est42:*
43: gb_est43:*

44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
 118: gb_est49:*
 119: gb_est50:*
 120: gb_est51:*
 121: gb_est52:*
 122: gb_est53:*
 123: gb_est54:*
 124: gb_est55:*
 125: gb_est56:*
 126: gb_est57:*
 127: gb_est58:*
 128: gb_est59:*
 129: gb_est60:*
 130: gb_est61:*
 131: gb_est62:*
 132: gb_est63:*
 133: gb_est64:*
 134: gb_est65:*
 135: gb_est66:*
 136: gb_est67:*
 137: gb_est68:*
 138: gb_est69:*
 139: gb_est70:*
 140: gb_est71:*
 141: gb_est72:*
 142: gb_est73:*
 143: gb_est74:*
 144: gb_est75:*
 145: gb_est76:*
 146: gb_est77:*
 147: gb_est78:*
 148: gb_est79:*
 149: gb_est80:*
 150: gb_est81:*
 151: gb_est82:*
 152: gb_est83:*
 153: gb_est84:*
 154: gb_est85:*
 155: gb_est86:*
 156: gb_est87:*
 157: gb_est88:*
 158: gb_est89:*
 159: gb_est90:*
 160: gb_est91:*
 161: gb_est92:*
 162: gb_est93:*
 163: gb_est94:*
 164: gb_est95:*
 165: gb_est96:*
 166: gb_est97:*
 167: gb_est98:*
 168: gb_est99:*
 169: gb_est100:*
 170: gb_est101:*
 171: gb_est102:*
 172: gb_est103:*
 173: gb_est104:*
 174: gb_est105:*
 175: gb_est106:*
 176: gb_est107:*
 177: gb_est108:*
 178: gb_est109:*
 179: gb_est110:*
 180: gb_est111:*
 181: gb_est112:*
 182: gb_est113:*
 183: gb_est114:*
 184: gb_est115:*
 185: gb_est116:*
 186: gb_est117:*
 187: gb_est118:*
 188: gb_est119:*
 189: gb_est120:*

Result No.	Score	Query Match	Length	ID	Description
1	963.8	37.6	977	105	AL514158
2	835.6	32.6	861	106	AL528484
3	636.8	24.9	692	113	AM299683
4	604.4	23.6	963	141	BE892079
5	596.6	23.3	767	18	AI316248
6	590	23.0	602	141	BE893726
7	581	22.7	620	111	AM152065
8	571.4	22.3	887	168	BE699109
9	523.6	20.4	541	122	AM966207
10	518.8	20.3	675	144	BE151259
11	515.6	20.1	815	145	AM044841
12	512.2	20.0	693	110	AM044841
13	506.2	19.8	889	141	BE914941
14	492.8	19.2	693	165	BE307752
15	485.4	19.0	618	174	BE245008
16	473.2	18.5	938	144	BE101858
17	470.4	18.4	487	5	AA315514
18	461	18.0	640	21	AI526513

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

19	449.4	17.5	462	5	AA305116
20	448.2	17.5	497	32	AV750182
21	446.8	17.4	616	32	AV714646
22	429	16.8	526	117	AW619441
23	409.4	16.0	544	8	AA475560
24	408.2	15.9	510	102	AI787419
25	404.2	15.8	624	144	BF138546
26	398.4	15.6	829	171	BF979524
27	386.8	15.1	437	149	BF509823
28	386.2	15.0	731	156	BE309622
29	384.6	15.0	482	165	BE309622
30	384.4	15.0	442	104	AI961078
31	383	15.0	538	8	AA475551
32	378.6	14.8	530	104	AI956534
33	374.4	14.6	484	157	W04752
34	368.4	14.4	527	32	AW012259
35	365.6	14.3	468	165	BE305849
36	356.4	13.9	466	3	AA213310
37	353.6	13.8	441	163	BE119891
38	350.6	13.7	366	9	AA622988
39	347.2	13.6	508	13	AA881807
40	346.8	13.5	474	14	AA986652
41	340.2	13.3	426	169	BF818728
42	339.6	13.3	483	113	AM259182
43	338.2	13.2	447	156	R89462
44	336.8	13.2	887	145	BF238534
45	334.2	13.0	625	166	BE372615

ALIGNMENTS

RESULT	1	LOCUS	AL514158	DEFINITION	AL514158 LFI_NFL006.PL2	mrna	EST	13-FEB-2001
ACCESSION	AL514158	AL514158	AL514158	Prime, mRNA sequence.	AL514158	GI:12777652		
VERSION	AL514158.1	GI:12777652						
KEYWORDS	EST.							
SOURCE	human.							
ORGANISM	Homo sapiens							
REFERENCE	1 (bases 1 to 977)							
AUTHORS	Li, W.-B., Gruber, C., Jessee, J. and Polayes, D.							
TITLE	Full-length cDNA libraries and normalization							
JOURNAL	Unpublished (2001)							
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.							

FEATURES
Source
1. 977
Location/Qualifiers

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Cl0BB0052B08"
/clone_id="LFI_NFL006.PL2"
/issue_type="placenta"
/note="Vector: pCMVSPORT 6; site_1: NotI; 1st strand cDNA was primed with a NotI-Oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@life.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 321 a 188 c 213 g 253 t 2 others

Query Match	37.6%	Score 963.8	DB 105	Length 977
Best Local Similarity	99.7%	Pred. No. 3.3e-156		
Matches 974	Conservative	2	Mismatches 0	Indels 1
				Gaps 1
Qy	738	cacctccctgatactacgaagcatggttaccacacccgggaagctctgcagatgtaca	797	
Db	1	CACCTCCCTGATCTACGAAGCATGTTTACCACACCGGAGGCTGCGAGATGTACA	60	
Qy	798	gtttatcacagcagtcgaagccgtgcatctggcaaaagctatctgaaagcaag	857	
Db	61	GTTTTATCACAGCTGCAAGCCGTGCGATTTGCAAGCATTTGATTTGAAGCAGCAAG	120	
Qy	858	gatggagaaatattgttatctgctgcgcaaaagccgcccacatccaaactctggc	917	
Db	121	GATGAGCAAAATATTGTTATTGTCGCAAAAGCCGCCACGACATCCAAAACCTTGAGC	180	
Qy	918	acaatctactgctgcgcaagaaattgaaacagcttggaagaaagcccttcatt	977	
Db	181	ACAATCTACTGCTGCTGCAAGAAATTCACACTTGGAGAAAGCCCTTGCAATTT	240	
Qy	978	gtgatgtgagagatgcaagcagatcagtgctgcagctggaagaaagccatcaagaatt	1037	
Db	241	GTTGATGTGAGAGATGGAACAGCATGCTGCTGAGTGGAGAAAGCCATCAAGAAATTT	300	
Qy	1038	ggaaggaattgatatctggttaataatgagcagctgcattagttgaccaataattggac	1097	
Db	301	GGAGGAATGTGATTCGTGTAATTAATGCGAGTCCCATTTATTTGACCAATATTTGAC	360	
Qy	1098	acacctaccagagatgagatcagatgagcgtggaagcgtgcaagcagagcacttacc	1157	
Db	361	ACACCTACCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG	420	
Qy	1158	tctaaagcatgattccttatttgaaagagcaagctgctcatatccatcaatacagt	1217	
Db	421	TCTAAAGCATGATTCCTTATTGAAAGAGCAAGTGTCTCATTCCTCAATATCAGT	480	
Qy	1218	ccaccactgaaactaaatccagttggttcaaaagcagctggttataccattgctaag	1277	
Db	481	CCACCACCTGAACTTAATCCAGTTTGTTCNAACGCACTGTGTTATTCACATTCGTAAG	540	
Qy	1278	tatgatatctatgatatgctctggaatgagcagaagaatttaagtgaaattgcaagtc	1337	
Db	541	TATGATATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATG	600	
Qy	1338	aatgcatatgagccttaaaacagcacaacacagctgctgatatgatatgctggagac	1397	
Db	601	AATGCAATATGCTGCTTAATAACAGCATACACACTGCTGATGATGATGCTGGAGACT	660	
Qy	1398	ggtatgaaagcagtgtagaagaattgatatcatctgcaagtgagatattccatttc	1457	
Db	661	GGTATCGAAGCCAGTGTAGAAAGTTGATATTCATTCGATGATGATGATGATGATG	720	
Qy	1458	caaaagccaaagcttctacgcaacttgctcatgtagtgaataatcttaaaagaa	1517	
Db	721	CAAAAGCCAAAGCTTTTACGCAACTTGTCTATGATGAATATCTTAAAGAAAGAA	780	
Qy	1518	ggaatagaanaatttgaagcttataatgaacacaggtcaatcccttgcaacagattc	1577	
Db	781	GGAATAGAAATTTTGAAGCTTATGCAATTAACACAGTCTCTCTTGCAACAGATTTC	840	
Qy	1578	ttctaatgataaccagaaagcagttgacgaagaagtggataacttgtagctgttcca	1637	
Db	841	TTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	900	
Qy	1638	gaattcaaaagaaagaaactgcaagctgcaacaaacacagcttgg-agctgtggaaga	1696	
Db	901	GAAATTCAAGAGAGAAAGTGCACCTGCAACCAAAACACAGCTTGTGATGCTGGAAGA	960	
Qy	1697	aacattagaattgtta 1713		
Db	961	DACATTTAGAAATGCTTA 977		

```

RESULT 2
AL528484      861 bp      mRNA      EST      13-FEB-2001
LOCUS      AL528484 LTI_NFL003_NBC3 Homo sapiens cDNA clone CS0DC029YE05 5
DEFINITION prime, mRNA sequence.
ACCESSION AL528484
VERSION AL528484.1 GI:12791977
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
    source
        1..861
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="CS0DC029YE05"
            /clone_1lb="LTI_NFL003_NBC3"
            /sex="male"
            /tissue_type="neuroblastoma cells"
            /lab_host="DH10B"
            /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
            was primed with a NotI-oligo(dT) primer. Five prime end
            enriched, double-stranded cDNA was digested with Not I and
            cloned into the Not I and Eco RV sites of the pCMVSPORT 6
            vector. Library was normalized. Library was constructed
            by Life Technologies. Contact : Feng Liang Life
            Technologies, a division of Invitrogen 9800 Medical Center
            Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
            8371 Email : fliang@lifestech.com URL :
            http://fulllength.invitrogen.com"
BASE COUNT 270 a 172 c 188 g 226 t 5 others
ORIGIN
Query Match 32.6%; Score 835.6; DB 106; Length 861;
Best Local Similarity 99.4%; Pred. No. 3.8e-134;
Matches 847; Conservative 2; Mismatches 1; Indels 2; Gaps 1;
QY 714 gctcgccgcgcgcgcgtctgtccgcgcacccctcctgtactacg--aaagtcagtgtaccga 771
    |||||||
DB 10 gctcgccgcgcgcgcgtctgtccgcgcacccctcctgtactacg--aaagtcagtgtaccga 69
QY 772 aacacgggagcgtgagcagatgtacagttttatcatcaaggtgcaagcgtgagcagtcgca 831
    |||||||
DB 70 acacgggagcgtgagcagatgtacagttttatcatcaaggtgcaagcgtgagcagtcgca 129
QY 832 aagcatctgcatgaaagcagcaagatgagcaaatctgtatctgtcgtcgaagaccg 891
    |||||||
DB 130 aagcatctgcatgaaagcagcaagatgagcaaatctgtatctgtcgtcgaagaccg 189
QY 892 cccagccacatccaaaactctcagcacaactatctactgtctgtcgaagaattggaagcg 951
    |||||||
DB 190 cccagccacatccaaaactctcagcacaactatctactgtctgtcgaagaattggaagcg 249
QY 952 tctggaagaaagcctctgcatgtatctgtcgtgagagatgaacagcagcagtcgctg 1011
    |||||||
DB 250 ttggagagaaagcctctgcatgtatctgtcgtgagagatgaacagcagcagtcgctg 309
QY 1012 cagtgagagaaagcagcatcaagaaatttgaggaattgatactctgtaaaatctgcaatg 1071
    |||||||
DB 310 cagtgagagaaagcagcatcaagaaatttgaggaattgatactctgtaaaatctgcaatg 369
QY 1072 ccatagtgtagccaaatataattgagacacccctcaagaagatgtggtctgtagatgaacg 1131
    |||||||

```

```

DB 370 CCATTAGTTTGACCAATACATTGGACACACTACCAAGAGATTGGATCTGATGTAACG 429
QY 1132 tgaacacagagagacacctaactctgcatctcaagaagcatgtatctcttatttgaaagagca 1191
    |||||||
DB 430 TGACACACAGAGGACACTTACCTTGATCTTAAGACATGTAATTCCTTATTTGAAAGACGA 489
QY 1192 aagtgctcatatcctcaatcagtcacacacacacacacacacacacacacacacacacac 1251
    |||||||
DB 490 AAGTTCCTATATCCCTCAATATCAGTCACACACTGAACCTAAATCCAGTTGGTTCAAC 549
QY 1252 agacagctgcttatcattacattgctgaagtatgctatgctctgtatgctctggaagtcag 1311
    |||||||
DB 550 AGCACTGCTCTTATACCATTTGCTAAGTATGTAATGTAATGTAATGTAATGTAATGTAATG 609
QY 1312 aagaatttaagtgaaatctgcaatgcatatgcatatgcatatgcatatgcatatgcatatg 1371
    |||||||
DB 610 AAGAAATTAAGGTGAATTTGCAAGTCAATGCAATTTATGCTTAAACAGCCATTAACACTG 669
QY 1372 ctgctatgatatgctggaagagacctgcatcgaagcagcagcagcagcagcagcagcagc 1431
    |||||||
DB 670 CTGCTATGATATGCTGAGGAGGACCTGATGCAAGCCAGTGTAAGAAAGTTGATATCA 729
QY 1432 ttgcagatgacgacatatctcatttccaaaagccaaaagtttactggaactgtgca 1491
    |||||||
DB 730 TTGCAGATGCAGCATATTCATTTTCCAAAGCCAAAGCTTTTACTGCAACTTTGTCA 789
QY 1492 ttgatgaaataatcttaaaagaagaagaatgaaatcttgacgtttatgcaatlaaac 1551
    |||||||
DB 790 TTGATGAAATAATCTTAAGAAAGAAAGAAATTAAGAAATTTGACGTTTATGCAATTAAC 849
QY 1552 cagtgatcctct 1563
    |||||||
DB 850 CAGGTATCTCTT 861
RESULT 3
AM299683/C 692 bp mRNA EST 18-JAN-2000
LOCUS xs42e02.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2772314 3'
DEFINITION similar to TR:O19066 O19066 OVARIAN STEROL CARRIER PROTEIN 2 ;,
            mRNA sequence.
ACCESSION AM299683
VERSION AM299683.1 GI:6709360
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 692)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel.: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmerit-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www.bio.llnl.gov/bbrp/image/image.html
Seq primer: -40up from Glbco
High quality sequence stop: 466.
FEATURES
    source
        1..692
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2772314"
            /clone_1lb="NCI_CGAP_Kid11"
            /lab_host="DH10B"

```


QY	1734	gtttttt-aaagcaatcaaacatctatctcg-cttgaactccgggtgaagtgtgtgca	1791
Db	481	gttttttAAAGCCACATCAACACATCTATCTGTTTAACTCCGGTGAATGGTGGCA	540
QY	1792	ctgtgtttcttcgatctgaaagaaggtgtggaaatgtcgagatagagagcctctgac	1851
Db	541	CGTGTGTTCTTGATCTGAAACAAAGGAGGGGCGCAGTCTCGAATGACAGAGCCTTCTGAT	600
QY	1852	agg---cagaatgtgtgatagtga-tyactactgactctgtataaaatgtcttcagg	1907
Db	601	CAGGCCAGATGTGTGTATAGTATGACTACTAGATGACTTTGTTAACTCGTTTC--GG	658
QY	1908	aaactaaaccaaacaatgycatctatctatgtcaaggaaatgtgaagt--aaagtataatcg	1965
Db	659	GAACTTAAACCAACAATGCGATCTCTGTCAAGGACCCCTCGAGATCACAGTAACATG	718
QY	1966	ccctagaacatcaaatctggagaagctatga	1995
Db	719	GCTTAGCAGCTCACTGTGGGAGAAAGTACAT	748

LOCUS	AI316248	767 bp	mRNA	EST	17-DEC-1998
DEFINITION	u726904.y1 Sugano mouse kidney mKia Mus musculus cDNA clone				
	IMAGE:1921110 5' similar to WP:C45B11.3 CE05427 ALCOHOL				
	DEHYDROGENASE ;, mRNA sequence.				
	DE112000				

ACCESSION	AI316248
VERSION	AI316248.1
FEATURES	GI:4031515

ORGANISMS

REFERENCE	AUTHORS
1 (bases 1 to 767)	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., . . .

TITLE	JOURNAL	COMMENT
The Washu-HHMT Mouse EST Project	Unpublished (1996)	Contact: Marra M/Mouse EST Project

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MGT:977402

Seq primer: custom primer used
High quality sequence stop: 483.

FEATURES	SOURCE
1. High Accuracy: The model achieves a high accuracy rate, consistently performing well across various datasets and tasks.	Model Performance Metrics
2. Scalability: The system is designed to handle large volumes of data and scale up to meet increasing demands.	System Architecture
3. Real-time Processing: The model is capable of processing data in real-time, enabling immediate decision-making.	Latency and Throughput
4. Interpretability: The model's decisions are transparent and explainable, allowing users to understand the underlying logic.	Model Explainability
5. Robustness: The system is resilient to adversarial attacks and maintains performance in noisy or incomplete data environments.	Security and Reliability
6. Integration: The model seamlessly integrates with existing infrastructure and third-party services.	System Compatibility
7. Customization: The system can be tailored to meet specific requirements and adapt to changing needs.	Flexibility and Adaptability
8. Cost Efficiency: The solution is optimized for cost, providing high performance at a lower price point.	Cost Analysis
9. Support: Comprehensive documentation and customer support are provided to ensure successful implementation.	User Support and Training
10. Future-Proofing: The system is designed with future updates and enhancements in mind, ensuring long-term relevance.	Product Roadmap

```

1. 767
/organism="Mus_musculus"
/strain="C57BL"
/db_xref="taxon:10090"
/clone="IMAGE:192110"
/clone_1lb="Sugano mouse kidney mk1a"
/sex="female"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: Kidney; Vector: pME18S-FL3; Site_1: DraIII
(CACGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
was primed with an oligo(dT) primer
[ATGTGGCTTTTCTTTTCTTTT]; double-stranded cDNA was
ligated to a DraIII adaptor [CTGGGCGCATGTG], digested
and cloned into distinct DraIII sites of the pME18S-FL3
vector (5' site CACGTGTG, 3' site CACCATGTG). XhoI should
be used to isolate the cDNA insert. Size selection was

```

performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTTCGCTCTAAAGCTCG and 3' end primer GCACCTGCAGCTCAGCACA."

Query Match	23.3%	Score	596.6;	DB	18;	Length	767;
Best Local Similarity	87.0%	Pred. No.	4.8e-93;				
Matches	667;	Conservative	0;	Mismatches	99;	Indels	1;
						Gaps	1

Qy 760 tcaagttacccaacaccggyagcgtgcgaagatctacagtlttcatcacagtgcaagcc 819
 ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | ||||| | |||||
Db 1 TCATGCTGCCCAACACTGGAGAGCTAAGCGAGATGCACAGTTTTCACACAGGTGCAAGCC 60

qy 820 ggcgcatcggcaaaagctatcgatcgtgaagcagaacaaaggatcgagcaaatctgcttatcg 8/9
+ + + + +
Db 61 GAGGCATTGGCGAAGCTATTCCTTGAAGAAGCCGATAGCATGGACCAATTTGTCTATTG 120

Db 121 CTGGGAGACCCACCAAAAGCACCCGAACCTCTCGGCACATCTACACGGATGCTGAAG 180

181 A A A T T G A A C G A C C T G G A G G G A C G C C T T G C C T T G T G T G A T G T G A G A G A T G A A C A G C 240

Db 241 ATATCACAAGTGCACGTGAGAAAGCTGTGGAGAAATTTGGAGGAATTGTTGGTGA 300

Db 301 ACAATGCCAGTGTCTATTAGCTTGACCAACAGCTTGAGACTCCGACCAAGAGAGTGAGACT 360

Db 361 TGATGATGATGTGAACACCGGGGCACCTACTTACATCCAAAGCATGTATTCTCTTTT 420

Db	421	TAAAAAAGAGCAAGTAGGTCATATTCTCAATCTCAGCCACCCCTGAACCTAAACCCAC	480
Qy	1240	tttgggtcacaacagcaactgtgcttatatcccatctgaagatgtaagtgtcatgtatgtatgc	1299

Db 481 TGAGGTTCAACACAGCAGCTGTCTTATACCATTCGCCAAATATGGCATGTCTATGTCTGTGC 540

QY 1300 ttggaatggcagaagaatttlaaagttgaatattgcagtcatagtcatattatggcctaaanacg 135

DB - 541 TTGGGATGGCTGAGACATTTTACAGGTCGAATTCAGTCATATGCCCTTATGGCCTAGTACAG 600
QY 1360 ccatacacactgctgatatgatatgctgagagacctggtatcgaaagccagtgagaa 141

Db 601 CCATTCACACTGCTCTATGCATATCGTGGGAGGATCTGGTGTGGAAACCAATGTAGAA 660
 Qy 1420 aagttgatatcatctgcagatgcagcatattccaatttccaaaagccaaaagaatttactg 1477

DD 001 ATGCTGACATCAATGCGACCGCATATATCCATTTTCATTAATGGCCATTAAGTTTACTG / 20
 Oy 1480 gcaacttgcattgtagaata-tcttaagaagaagaataga 1525

RESULT 6

LOCUS	602 bp	mrna	EST	20-OCT-2000
DEFINITION	BE893726	601436638F1	NIH_MGC_72	Homo sapiens CDNA clone IMAGE:3921920 5'
	mrna sequence.			

ACCESSION	BE893726
VERSION	BE893726.1
	GI:10355379

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 602)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: ATCC/DCPD/DRP
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov
plate: L1AM9755 row: k column: 09
High quality sequence stop: 602.
Location/Qualifiers
1. 602
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3921920"
/clone_lib="NIH-MGC_72"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pCMV-SPORT6; Site: 1: NotI; Site: 2: SalI; Cloned unidirectionally. Primer: oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."
BASE COUNT 196 a 109 c 129 g 168 t
ORIGIN

Query Match 23.0%; Score 590; DB 141; Length 602;
Best Local Similarity 99.8%; Pred. No. 6.6e-92;
Matches 601; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 1209 aatacagtcacacact-gaacctaaatccagttggttcacaacagcactgtgtctatc 1267
|||||
DB 1 AATATCACTCCACCCGTCGAACCTTAATCCAGTTGGTTCAACAGCACTGGCTTATAC 60
|||||
OY 1268 catgtcctaagatggtatgctatgctatgcttggaatgagcagaagaattaaagtga 1327
|||||
DB 61 CATTGCTAAGATGATGTATGTATGTATGTCTGGAATGCGAAGAATTTAAAGTGA 120
|||||
OY 1328 aattgcacatcatatgacctaaacagccatcacacactgtgtctatgatatgt 1387
|||||
DB 121 AATTGCAGTCAATGATATGCTTAACCAACCCATACACTGCTGCTATGATATGCT 180
|||||
OY 1388 gggaggaactgtgtacgaagccagttgtagaagaatgatcatctgcagatgcagaca 1447
|||||
DB 181 GGGAGGACCTGGTATCGAAGCCAGTGTAGAAAGTTGATATTCAGATGCAGCATTA 240
|||||
OY 1448 ttcaatttccaaaagccaaaagtcttactgagcaactgttcattgtaaaatactt 1507
|||||
DB 241 TTCCATTTTCCAAAAGCCAAAAGTTTACTGCGCAACTTTGTCATGTGATAAATATCTT 300
|||||
OY 1508 aaaagaagaagaatagaatatttgagcgttatgcaatlaaacaggtcatcctttga 1567
|||||
DB 301 AAAAGACAGCAAGATAGAAAATTTTGACGTTTATGCAATTAACCGAGTCACTCTTCA 360
|||||
OY 1568 accagattcttctagaatgataccagaaagcagttgagaagaagtgaatcaactgg 1627
|||||
DB 361 ACCGATTTCTTTAGATGAATACCCAGAGCATTTGCAAGAAAGTGATCACTACCTGG 420
|||||
OY 1628 tgcgtctcagaatcagaagaagaacatgcagctgcaaccaaaacacagtlctgagac 1687
|||||
DB 421 TGTGTTCCAGAAATTCAGAAAGAACTGACAGCTGCAACCAAAAACACGCTTGGAGC 480
|||||

OY 1688 tgtggaagaacatttaaatgttaagactctcagtgatgattgtttaagccac 1747
|||||
DB 491 TGTGGAAGAACATTTAAGATTTTGAAGACCTCTCAGTATGATGTTGTTAAAGCCAC 540
|||||
OY 1748 tcaagcaatctatcgtttgactctccggtgaagatggtgcaagttcttcatc 1807
|||||
DB 541 TCAACCATCTATCTGTTTGAACCTCCGCGTAACATGCTGCGACGCTGTTCTGATCT 600
|||||
OY 1808 ga 1809
||
DB 601 GA 602
|||||

RESULT 7
AM152065/c 620 bp mRNA EST 03-NOV-1999
LOCUS x174a01.x1 NCI-CGAP-Gas4 Homo sapiens CDNA clone IMAGE:2623752 3'
DEFINITION similar to WP:CI7G10.8 CE16861 FAT-3: ALCOHOL DEHYDROGENASE ;, mRNA
sequence.
ACCESSION AM152065
VERSION AM152065.1 GI:6199963
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 620)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www-bio.llnl.gov/bdrp/image/image.html
Seq primer: -40UP from Glibco
High quality sequence stop: 416.
Location/Qualifiers
1. 620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2623752"
/clone_lib="NCI CGAP Gas4"
/tissue_type="poorly differentiated adenocarcinoma with signet ring cell features"
/note="Organ: stomach; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: oligo dT. Average insert size 1.69 kb. Life Technologies catalog #: 11549-011"

BASE COUNT 162 a 144 c 101 g 209 t 4 others
ORIGIN

Query Match 22.7%; Score 581; DB 111; Length 620;
Best Local Similarity 98.2%; Pred. No. 2.3e-90;
Matches 606; Conservative 0; Mismatches 9; Indels 2; Gaps 2;

OY 1404 gaaagcaggtgtagaagaatgatatcatctcagatgcagcatcttccatccaaag 1463
|||||
DB 620 GAAGCCCAAGTGTAGAAAGTTGATATCATGCGAAGTGCAGCATATTTCCAAAAA 561
|||||
OY 1464 ccaaaaag-ttctaagcgaacttgtt-cattgatgaataatcttaaaagaagaaggaa 1521
|||||
DB 560 GCAAAAAGTTTACTGCAACTNCTGTCATGTGATGAATAATATCTTAAGAAGAGCAA 501
|||||

```
QY 1522 tagaataattgcagctttagcaatgaacacagatcatccttgcacacagattctct 1581
|||||
Db 500 TAGANATTGTCGCTTATGCAATTAACCAAGGTCATCCTTGCACAGATTTCTCT 441
QY 1582 tagatgaataaccagaagcagtttagcaagaagaattggaatcaactgtgtcttccgaat 1641
|||||
Db 440 TAGATGAAATACCCAGAGAGAGATGACAGAAAGTGAAATCACTGCTGTCTCCAGAAAT 381
QY 1642 tcaaaagaagaacacagctgcacacaaacacacacacacacacacacacacacacacac 1701
|||||
Db 380 TCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 321
QY 1702 ttagaattgttaagaagctctcagatgtagatgtgtttaaagcactcaagaactatc 1761
|||||
Db 320 TTAGAATTGTTGAAGACTCTCTCAGATGATGATGTTGTTAAAGCCTCAAGCAATCTATC 261
QY 1762 tgttgaactctccagtgagaatgtgtgagcagtggttcttgaatcgaagaagaaggtg 1821
|||||
Db 260 TGTGTGAATCTCTCGGTGAGAGATGAGTGCACGTTCTTGAATCTGAAGAGAGAGAGAG 201
QY 1822 ggaatgcgagatagagaagcctctcagacagcagatgtgtgtagatgtagatgtagatg 1881
|||||
Db 200 GGAATGTCGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 141
QY 1882 atgaactgttaaaatgcttcaaggagaacaaacaaacaaatgagcattcagtcagga 1941
|||||
Db 140 ATGACTTTGTAATAAATGTTTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 81
QY 1942 aattgaagatgaagttgaacatgagcctcagcagcaatcgaattggaagaagcgaatcaga 2001
|||||
Db 80 AATTGAAGATTAAGGTAACATGAGCCCTAGCAATCAATTTGAGAGAGAGAGAGAGAGAGAG 21
QY 2002 tgaatgcagcagactgtga 2018
|||||
Db 20 TGAATGCCAGACTGTGA 4

RESULT 8
BF699109 887 bp mRNA EST 22-DEC-2000
LOCUS 602126730F1 NIH_MGC_56 Homo sapiens cDNA clone IMAGE:4283671 5',
DEFINITION mRNA sequence.
ACCESSION BF699109.1 GI:11984517
VERSION BF699109.1
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 887)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.L.Strausberg@nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM1115 row: 1 column: 08
High quality sequence stop: 611.
Location/Qualifiers
1.887
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4283671"
/clone.lib="NIH_MGC_56"
/tissue.type="primitive neuroectoderm"
/lab_host="DH10B (T1 phage-resistant)"
```

```
/note="Organ: brain; Vector: pDNR-LIB (Clontech); Site:1:
SfiI (ggcgccctcgccg); Site:2: SfiI (ggcgcttaagcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-ATCTTAGAGCGCGGCGCGCGAGATC-dT(30)BN-3'
(sequence: 5'-ATCTTAGAGCGCGGCGCGCGAGATC-dT(30)BN-3'
where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
BASE COUNT 251 a 185 c 227 g 224 t
ORIGIN
Query Match 22.3%; Score 571.4; DB 168; Length 887;
Best Local Similarity 95.5%; Pred. No. 1e-88; 26; Indels 3; Gaps 3;
Matches 620; Conservative 0; Mismatches 26; Indels 3; Gaps 3;

QY 1114 tgaatctgaatgaacgctgaacacacagagacacacacacacacacacacacacacacac 1173
|||||
Db 12 TGGATCTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 71
QY 1174 cttaattgaaagaagcaagtgctcattatcattcattcattcattcattcattcattcattcatt 1233
|||||
Db 72 CTATTTGAAAAAGACCAAGTTGCTCATATCTCTCAATATCATGCTCCACCATGAACTTAA 131
QY 1234 atcgaactgtgttcaaacagacacacacacacacacacacacacacacacacacacacacac 1293
|||||
Db 132 ATCCAGTTTGGTTCAAAACAGCACTGCTTATACATTCCTTAAGTATGATGATGATGATG 191
QY 1294 atgtgtctggaatgagagagaatgaagaagtgaagaatgagagagagagagagagagagag 1353
|||||
Db 192 ATGCTGCTTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 251
QY 1354 aaacagacacacacacacacacacacacacacacacacacacacacacacacacacacacac 1413
|||||
Db 252 AATACAGCCATACACACTGCTGCTATGATATGCTGGAGAGAGAGAGAGAGAGAGAGAGAG 311
QY 1414 gtagaagaatgtgatalcattcagatgcagatgcagatgcagatgcagatgcagatgcagatgc 1473
|||||
Db 312 GTAGAAAGGTGATATCATTTGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 371
QY 1474 ttaactggaactgtgtcattatgataaataatcattaaagaagaagaagaagaagaagaatttg 1533
|||||
Db 372 TTACTGCAACTTGTTCATTTGATGATAATATCTTAAAGAGAGAGAGAGAGAGAGAGAGAG 431
QY 1534 acgttatcgaattaaacacagatcattccttgcacacagattcctcttagatgaattacc 1593
|||||
Db 432 ACCTTTATGCAATTAACCAAGGTCATCTTTGCAACCAAGTTTCTTATGATGATGATGATG 491
QY 1594 cagaagaagctgaagaag-aaagtgaatcaactgtgtctgttccagaatcagaagaagaaga 1652
|||||
Db 492 CACAAGCAGATTACAGAGAAAGTGAATCACTGAGTG-TCACAATTTCAAGAGAGAG 550
QY 1653 aaactgagctgacacc-aaacacacgcttcagagctgtggaagaagaacattagaattgt 1711
|||||
Db 551 AAATCTCAGCTGCAACCAAAACACAGCTTCTGAGCTGTGGAAGAAACATTTAATGTT 610
QY 1712 taagagctctcctcagtgatgattgtttaaagcactcaagaacattat 1760
|||||
Db 611 AGGACTCTCTCGTGGTATGATGATGCTTTAAGCCCTCAGCAACTATCTGT 659

RESULT 9
AW966207 541 bp mRNA EST 01-JUN-2000
LOCUS AW966207
DEFINITION EST178280 MAGE resequences, MAGI Homo sapiens cDNA, mRNA sequence.
ACCESSION AW966207
VERSION AW966207.1 GI:8156043
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```


Oy	1047	gattttcttgtaaaataatgcagtgccattagtgttaaccaaatacatatggagcaaacctc	1106
Dd	301	gattatttttggttaacacatccactgcgtattattgcttgaccacacggttgacactccgacc	360
Oy	1107	aagagattgacatcgtatgatacgttgaaacacacagagacacactacacttgatctaaagca	1166
Dd	361	aagagagtgctgacttgattgattgattggaacacacaggggaccttaccttacctcaaaagca	420
Oy	1167	tgatttccttatttggaaaaagacaaagttgctcatalactcctaatacagtcacacacgt	1226
Dd	421	tgtatttcccttttttttaaaaaagacaaagatgctcatatttctcatctcacaccatccctg	480
Oy	1227	aaccttaaacccagtttggttccaacagcactggtctataccattgctaagtatgatatg	1286
Dd	481	aaccttataccactctgggtttccaacacgactggtctttatnccatttggccatattggcgatg	540
Oy	1287	tctatgtatgctcttggaatggtgcagaagaatttaaaggttaaatggcagtcacatgcatla	1346
Dd	541	tctatttgctgctcttgggcatggctggaagaattnagaggttaaatttgcacgtcttatgccctta	600
Oy	1347	tgagcctaaacagccatacacactgctgctatgataatgctggagaggaaccttgatcgaa	1406
Dd	601	tgcgtctagaaacagtcatttccacactgcgtgcattgatattgcttgagagatctggcgctgat	660
Oy	1407	agccagtgctagaana 1421	
Dd	661	aatcatattgtagaaaa 675	

RESULT	11
LOCUS	Bf182478
DEFINITION	Bf182478 B15 bp mRNA EST 31-OCT-2000 601804245P1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4035179 5'
ACCESSION	Bf182478
VERSION	Bf182478.1 GI:11060621
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 815) NIH-MGC http://mgc.nci.nih.gov/ . National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
AUTHORS	Contact: Robert Strausberg, Ph.D.
JOURNAL	Tel.: (301) 496-1550
COMMENT	Email: Robert.Strausberg@nih.gov Tissue procurement: Lothar Hennighausen Ph.D., Robin Humphreys cDNA library preparation: Life Technologies, Inc. cDNA sequencing by: The I.M.A.G.E. Consortium (LLNL) DNA distribution: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLMW9308 row: j column: 12 High quality sequence stop:660.

BASE COUNT	FEATURES SOURCE
	Location/Qualifiers
	1. .815
	/organism="Mus musculus"
	/strain="C57/B6"
	/db_xref="taxon:10090"
	/clone="IMAGE:4035179"
	/clone_1lb="NCI_CGAP_Mam5"
	/tissue_type="tumor, gross tissue"
	/dev_stage="7 months"
	/lab_host="DH10B"
	/note="Organ: mammary; Vector: PCMV-SPORT6; Site:1: Saliv; Site:2: Notti; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Robin Humphreys, NIH"
250 a	156 c 211 g 198 t

ORIGIN

Query Match	20.1%;	Score 515.6;	DB 145;	Length 815;
Best Local Similarity	83.1%;	Pred. No. 4.1e-79;		
Matches 635;	Conservative 0;	Mismatches 124;	Indels 5;	Gaps 4;

QY	903	ccaaactctcgaagcacaactctaacgcgcgcgaagaaattgnaacaaattggagggaag	962
Db	1	ccgaacactccctcgccacacattctacacggcgcgtgaaanaattgaamccactctgagggacg	60
QY	963	gcctctgcacatgtaactgtctgatactgtagaagatgaaacagacagatcagctgctcagttgagaa	1022
Db	61	gccttgccttctgtttgtttgaattgtsagagattgaacacagaatactaacacgtctcactggagaaa	120
QY	1023	gcacatcaagaaatttggaggaattgatactctcgtaataataatgcccagctgacattggtt	1083
Db	121	gctgtggagaaattttggaggaattgatttttggctgaacacattcccgctcttattagcttg	180
QY	1083	accataataattggacacacacctaccagaagattgatactgatactgaatgaagtgaaccccga	1144
Db	181	accaaacacgcttggagacctccgacacaaagagtgactgtatgtatgtaattgaaacccacg	240
QY	1143	ggcaectacctctgatactctaaagcagtcatctcctatttgaanaagacaaagtctcat	1207
Db	241	ggcacttactcttatactcaaaagcatgatttctttttttaaanaagcaaaatagctcat	300
QY	1203	atccctcaatcagtcacacacactgaacctaatccagtttggctcaacagacactgtgct	1267
Db	301	attctcaattctcacaccacccctgaacctgaacccactgctgttcaaacacagacactgct	360
QY	1263	tataccattgttaagtatgtatgtctcattatgttgctgtgaaatggcagaanaatttaa	1322
Db	361	tattaccatttgcacaaatttggatgtctatgtgtgtgtggatggcttaaacaaattttga	420
QY	1323	ggtgaattgcagtcacatgcatctatgtgcctcaaaacagccatacacactgctcctagat	1383
Db	421	ggtgaattgcagtcacatgcttattgtgcccrtaaagaaacacatttcacactctctctatgat	480
QY	1383	atgctggaaggaacctggtatcgaagaagcagtgtagaaaagtgtgatactatgcatagca	1443
Db	481	atgctggaaggaacctggtgtttgaanaaccaaattgtagaaaattgtacattcattcgacacct	540
QY	1443	gcatactccattctccaaaagccaaaagtttcaactgcgcgaacttgctcttga--tgaa	1500
Db	541	g-atatttgcattttt--caaaaagcaaaaaggttttaccggcaactttttttatgaatgaana	598
QY	1501	atactctaaagaagaaggaatagaanaattgaactttagctatctgaactctaaacaggtctac	1560
Db	599	atattctttaaagaagaagaatcacaacaaattttgatctctacgcaattttgcaccagctcat	658
QY	1561	cttctgcaacaggaattctctcttgaatgaataccgaagaagcagttgaacaaagaatgtaat	1620
Db	659	cccttgcattacggaattttctttttagatganaaccagatgcagatgattggggagagaaaana	718
QY	1621	caactgcgtgcgtctccagaattcaaaaagaagaanaactgcaactg	1664
Db	719	cacatgat--ctgtgcccgaagtgtgagaacaaanaagcggcgccg	761

RESULT	12
AM044841	
LOCUS	
DEFINITION	
AM044841	693 bp mRNA
un14e01.v1	Sugano mouse kidney mRna Mus musculus cDNA clone
IMAGE:2192280.5	similar to WP:045B1.3 CE05427 ALCOHOL
DERIVEDFROM:BASE	;; mRNA sequence.
ACCESSION	
AM044841	
VERSION	
AM044841.1	GI:5905370
KEYWORDS	EST.
SOURCE	
ORGANISM	
house mouse.	
Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus	
18-SEP-1999	

REFERENCE	(bases 1 to 693)
AUTHORS	Marras,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE	The Mashu-NCI Mouse EST Project 1999
JOURNAL	Unpublished (1999)
COMMENT	Contact: Marra M/Mashu-NCI Mouse EST Project 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.edu This clone is available royalty-free through LINDA; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MG1:1004732
FEATURES	Seq primer: custom primer used High quality sequence stop: A54. Location/Qualifiers
SOURCE	1..693 /organism="Mus musculus" /strain="C57BL." /db_xref="taxon:10090" /clone_image="IMAGE:2192280" /clone_id="Sugano mouse kidney mk1a" /sex="female" /dev_stage="adult" /lab_host="DH10B" /note="Organ: kidney; Vector: pME185-FL3; Site_1: DraIII (CACGTGTC); Site_2: DraIII (CACCATGTGC); 1st strand cDNA was primed with an oligo (dt) primer (ATCGGCCCTTTTCTTTTTTTTTTTT); double-stranded cDNA was ligated into a DraIII adaptor (TGTTGGCGCTACTGG), digested vector (5' site CACGTGTC, 3' site CACCATGT). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTCGCTCTAAAGCTCG and 3' end primer CGACTCTCAGCTCGACACA."
BASE COUNT	193 a 156 c 175 g 105 t 4 others
ORIGIN	
Query Match	20.0%; Score 512.2; DB 110; Length 693;
Best Local Similarity	86.1%; Pred. No.1.6e-78;
Matches 576; Conservative 0; Mismatches 92; Indels 1; Gaps 1;	
OY	733 gccgcacacctcctcgtacatcaagaagcatgtaaccacaacccggaggc-tggcagga 791
DB	25 gccacacagttctcgaaattgcttattgttcatactctcccaacacttgamaacggcgcgga 84
OY	792 tgtacagttttatcaacagtgcaagccgtygcatitggcaaatgatctatgcatitgaaga 851
DB	85 TGCAcAGATTtTAfTAcACAGCTGCAAGCGAGCATTGGCAAAAGCAATTgCCtTGAAGACC 144
OY	852 gcaaaagtatgagcaaatatgttatgtctgctgaagaccgcccgacacatccaactt 911
DB	145 GCAAAAGATGAGCCCAATATTtTCTCTGTCGGAAGACCACCCAATAACCCAGAACCTC 204
OY	912 cttagcacaaatcatcactgcctgctaagaatitgaagcagtttgagaggaagcccttgcca 971
DB	205 CTCGGGCACAAATtTAACCGGCTCTCTAAGAATAATGAAGCAGCTGGAGGAGCGCCTTGCTT 264
OY	972 tgatatgttgatgtagagagatgaacagacagatcagtgctgcagatgtagagaaaagccatcaa 1031
DB	265 TGTGTGTGTGAAGTAGAGATGAACAGCANATCAACAGACAGTGCAGAGAAACCTGTGGAG 324
OY	1032 aaatttgaggaattgatatcttgtttaataatagccagatgcccattagtttgaccaataca 1091
DB	325 AAATTtTGAAGGAATTGATrATTtTGGTtTGAACAATGCCAGCTGCTATTtTAGCTTGCAACACG 384

QY	1092	ttggacacacctaccagaagattgga	tc	tgatgatgaacgctgaacacagagcactac	1151
Db	385	TTGGACACTCCGACCAAGAGAGTGACTTGATGATGATGATGACACACGAGGACACTTAC	444		
QY	1152	cttgatctaagaagctgattctctatttgga	aaagacaagtgtgcctatctccaat	1211	
Db	445	CTTAACTCCAAAGCATGATTCCTTTNTTAAAAAGACCAAGATGAGTCATATTCCTCAAT	504		
QY	1212	atcagtcacaccactaacctaatactcagtttggttca	aaagcaactgtgctataccatt	1271	
Db	505	CTCAGCCCAACCCCTTAACCTTAACCCACTGTGTGTTCAACACGCACTGTGTATATACCAT	564		
QY	1272	gctaaagtgtgatgtatcgtatgtgtcttgga	atgcaagaagaatttaaagtgtaaat	1331	
Db	565	GCCAAATATGCGATGTATGTGTGTGCTGGATGCGTGAAGAATTATAGAGTGANMTT	624		
QY	1332	gcagtcacatgcatatagcctaaacagacatacacactgtcgtatgatatgctgtgga	1391		
Db	625	GCAGTCATATGCGCTTATGCGCTTAGAACACAGNCATTCACACTGCTGTATGATATGCTGGGA	684		
QY	1392	ggacctggt	1400		
Db	685	GGATCTGGT	693		
RESULT	13				
LOCUS	BE914941	889 bp	mRNA	EST	29-SEP-2000
DEFINITION	601667975F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3967867	5'			
ACCESSION	BE914941				
VERSION	BE914941.1	GI:10414077			
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
TITLE	1 (bases 1 to 889)				
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Tel: (301) 496-1550				
	Email: Robert.Strausberg@nih.gov				
	Tissue Procurement: Gilbert Smith, Ph.D.				
	cDNA Library Preparation: Life Technologies, Inc.				
	DNA Sequencing by: Incyte Genomics, Inc.				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LLNL at:				
	http://image.llnl.gov				
	Plate: LLAM9143 row: e column: 20				
FEATURES	High quality sequence stop: 632.				
source	Location/Qualifiers				
	1..889				
	/organism="Mus musculus"				
	/strain="FVB/N"				
	/db_xref="taxon:10090"				
	/clone_image="IMAGE:3967867"				
	/clone_id="NCI_CGAP_Mam1"				
	/tissue_type="tumor, biopsy sample"				
	/dev_stage="3 months, virgin"				
	/lab_host="DH10B"				
	/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;				
	Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.				
	Library constructed by Life Technologies. Investigator				
	providing samples: Gilbert Smith, NIH"				
BASE COUNT	246 a 199 c 224 g 220 t				
ORIGIN					

QY 1343 attatgacctaaacagccatcacactgctct-atgagatctgctggagagactgcta 1401
 DB 597 cttatgggct-AAACAGCCATTACACCTGCTGTATGATGATGCGAGATCTGTCG 655
 QY 1402 tcgaaagccagctgtagaaagctgatatcatctgc 1435
 DB 656 ttgacacaccattgtagAACACGTTGACATCATTCG 689

RESULT 15
 BG245008 618 bp mRNA EST 13-FEB-2001
 LOCUS BG245008 602358280F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:4486915 5',
 DEFINITION mRNA sequence.
 ACCESSION BG245008.1 GI:12754823
 VERSION BG245008
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 618)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC Clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10330 row: P column: 20
 High quality sequence stop: 618.

FEATURES
 source location/Qualifiers
 1..618
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4486915"
 /clone_lib="NCI_CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="3 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: PCMV-SPORT6; Site:1: Salt;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 183 a 134 c 156 g 145 t
 ORIGIN

Query Match 19.0%; Score 485.4; DB 174; Length 618;
 Best Local Similarity 87.6%; Pred. No. 6.5e-74;
 Matches 542; Conservative 0; Mismatches 76; Indels 1; Gaps 1;

QY 773 caccgggaagcgtgagagatgacagttttatcaaggtgcaagccgtgagcattggcaa 832
 DB 1 CACTGGGAAGCTAGCGGAGTGCACAGTTTTCACAGGTGCAGCGCGCATTTGGCNA 60

QY 833 agctattgcatgaaagcagcaagagatgagcaaatatgttatgtctgcaaaagaccgc 892
 DB 61 AGCAATTGGCTTGAAGCGCGCAAGAGATGAGCCATATTGTCATTGCTGCGAAGCAC 120

QY 893 ccagccacatcccaaaactctgagcaaatctatactgctgctggaagaaattggaagcag 952
 DB 121 CCAAAAGCACCCGAAATCTCTGCGCAACATCTACAGCGCTGCTGAAGAAATTGAAGCAC 180

QY 953 tggaggaagagccttgcatgtatgttgatgagatgaaacagcagatcagtgctgc 1012

DB 181 tggaggaagcagccttgctcccttgcg-ctgttatgtgagacagtgaacacaaatcaacagctgc 239
 QY 1013 agtggagaagccatcaagaanaattggaggaattgatalctctgtaataatgagcagtcg 1072
 DB 240 AGTGGAGAAAGCTGTGAGCAAAATTTGGAGGAATTGATTTATTTGTTGACAAATGCGAGTGC 299
 QY 1073 catgttgtagcaaatatgtagacacactcccaagagatgttgatctgtagatgaagt 1132
 DB 300 TATTAGCTTGGACCAACACGTTGGACACCTCGACAGAGAGAGAGACTTGATGATGAATGT 359

QY 1133 gaacacaaaggaagcacttacccttgatcaatgaagcatgtatcccttatttgaagaagagca 1192
 DB 360 GAACACAGGGGCGACCTTACCTTACATCCAAAGCATGTATTCTTTTAAAAAGAGCA 419

QY 1193 agtgcatactcctcaataatcagtcacacactgaaacctaaatccagtttgltcaaca 1252
 DB 420 AGTAGGTCAATATTCATCAATCTCAGCCACCCTGAACCTTAACCCACTGTGTTCAACA 479

QY 1253 gcaactgtcttatccatgctgaagatgtagtctatgtatgttgctggaatggcaga 1312
 DB 480 GCACCTGCTTATACCATTTGCCAAATATGGCATGTATGTGTGCTTGGAGTGGCTGA 539

QY 1313 agaattlaaaggtgaaattgcagtcacatgcatatgacctaaacagccatcacactgc 1372
 DB 540 AGAATCTAGAGTGGAATTTGCAAGTCAATGCTTATGCGCTTAGAACAGCATTCACACTGC 599

QY 1373 tgcatagtatgctggga 1391
 DB 600 TCTATGATGATATGCTGGGA 618

Search completed: June 12, 2001, 10:40:35
 Job time: 4553 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 09:59:37 ; Search time 89.72 Seconds
(without alignments)
4985.922 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561

Sequence: 1 agcagcagaatgatgcaaacga.....gsgmggraswmwawrrammc 2561

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/6D.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	556.2	21.7	4951	2	US-08-752-307B-1
2	400	15.6	633	1	US-08-644-664B-6
3	400	15.6	633	2	US-08-761-277A-6
4	259.4	10.1	9045	3	US-09-121-321-1
5	249.4	9.7	720	1	US-08-217-210B-3
6	181.4	7.1	422	1	US-07-920-519-33
7	181.4	7.1	422	1	US-08-314-586-33
8	181.4	7.1	422	1	US-08-314-121-13
9	181.4	7.1	422	2	US-07-927-391-14
10	167.6	6.5	7287	2	US-08-659-206A-1
11	166	6.5	6253	3	US-08-893-327-15
12	166	6.5	6280	3	US-08-893-327-17
13	166	6.5	6280	3	US-08-893-327-19
14	166	6.5	6295	2	US-08-659-206A-4
15	166	6.5	7852	2	US-08-836-022A-2
16	166	6.5	7852	4	US-09-427-048A-2
17	166	6.5	7892	2	US-07-916-098A-40
18	166	6.5	7897	4	US-08-836-022A-1
19	166	6.5	7897	4	US-09-427-048A-1
20	166	6.5	8236	1	US-08-461-837-1
21	166	6.5	8236	5	US-08-973-223-1
22	166	6.5	8236	5	PCT-US96-09495-1
23	166	6.5	8509	1	US-08-462-014-1
24	166	6.5	10398	2	US-08-331-384-1
25	166	6.5	10398	2	US-08-708-188-1
26	166	6.5	10398	2	US-08-836-087-1
27	166	6.5	19307	3	US-08-836-022A-10

C 28	166	6.5	19307	4	US-09-427-048A-10	Sequence 10, Appl
C 29	163.6	6.4	7286	4	US-09-331-581-3	Sequence 3, Appl
C 30	163.6	6.4	7938	4	US-09-331-581-14	Sequence 14, Appl
C 31	163	6.4	5141	1	US-08-286-305A-9	Sequence 9, Appl
C 32	163	6.4	5141	2	US-08-441-104A-9	Sequence 9, Appl
C 33	163	6.4	5141	2	US-08-440-816A-9	Sequence 9, Appl
C 34	163	6.4	6557	1	US-08-286-740-3	Sequence 3, Appl
C 35	163	6.4	6557	5	PCT-US95-09576-3	Sequence 3, Appl
C 36	163	6.4	8120	3	US-09-027-449-68	Sequence 68, Appl
C 37	163	6.4	8120	4	US-09-026-985-68	Sequence 68, Appl
C 38	161.8	6.3	10596	1	US-07-884-811-15	Sequence 15, Appl
C 39	161.8	6.3	10596	1	US-07-885-971-15	Sequence 15, Appl
C 40	161.8	6.3	10596	1	US-08-087-783A-15	Sequence 15, Appl
C 41	161.8	6.3	10596	1	US-08-194-088B-15	Sequence 15, Appl
C 42	161.8	6.3	10596	2	US-08-194-087-15	Sequence 15, Appl
C 43	161.8	6.3	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C 44	161.4	6.3	10785	3	US-08-444-644-27	Sequence 27, Appl
C 45	161.4	6.3	10844	3	US-08-444-644-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-08-752-307B-1
; Sequence 1, Application US/08752307B
; Patent No. 5952171
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; APPLICANT: Geating, David P.
; APPLICANT: Levinson, Douglas A.
; TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; City: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,307B
; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; TELEPHONE/DOCKET NUMBER: 09404/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4951 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-752-307B-1
Query Match 21.7%; Score 556.2; DB 2; Length 4951;
Best local Similarity 96.2%; Pred. No. 3,7e-105;
Matches 656; Conservative 0; Mismatches 14; Indels 12; Gaps 8;

```

QY      1  aggcgaagatgtgcgaagcagcgccttccaaattagtcagcaaacatagctccggccct 60
Db      131  AGCGCAAGATGTGCAGAACGATGCATCTTC -AATTAGTACAC -AACCATATGTCCTC -GCCCT 187
QY      61  aactcgcacatcccgccctaactccgncacagttccgycacatctccgcccatac 120
Db      188  AACCTCGCCCATTCGCCGCCCTTACCTCCGCC - -AGTTCGCCCATCTCTCGGCCCATGCGC 245
QY      121  tgacctaattttttattatagaaga -gcagagcgcgcttcgaccttgagctattccag 179
Db      246  TGACCTAATTTTTTTTATTTATGACAGAGGCCAGAGCCGCCCTCGGCTCTGAGCTATTTCCAG 305
QY      180  aagtagtaggaaggccttttttgagggccctagagcttttgcaaaaagccctc -cgaicag 238
Db      306  AAGTAGTAGGAGAGGCTTTTTTGAGAGCCTTAGCTTTTGCAAAAAGCTCTCTCCATTCAGC 365
QY      239  ggctcgcaatctctccctccacagcgccgcgcctactagtagccgcacatcaagcggt 298
Db      366  GGCTCGCATCTCTCTTCACCGCCGCCGCCCTTACCTGAGAGCCGCATCACGCGCGTT 425
QY      299  gaatcgagctttgcgcgcctcccgagcttbgcgctcctgaactagcgctcagtgag 358
Db      426  GAGTGCCTTCTGCGGCTCTCCGCTGtGtGtCCTCTGAACTGCTGCCCGCTTAGGT 485
QY      359  aagttlaaagctcagtcagtagaccgggacctttgfcgagcgctccctcttgagcctac 418
Db      486  AAGTTTAAAGCTCAGAGTCGAGACCGGGGCTTTGTCGCGCGCTGCTTGAGAGCTACTAG 545
QY      419  actaagcgagctctccacagctttgctcgacccctgcttgcactaactagctcttgct 478
Db      546  ACTCAGCGCGGCTCTCCAGCGCTTTCCTGACGCCCTTCGCTCAACTCTACCGCTTGTTC 605
QY      479  agttctcgtctcgcgcggttacagatccaaagctcgaaaaacagaagttiaactgta 538
Db      606  -GTTTTCGTTCTCGGCCGTTACAGATCCAAAGCTCTGAAAAACAGAAATTAACTGTGA 664
QY      539  agtttagtcttttgcctttatttcagttccgagatccg - -tggtygtycaaatca 594
Db      665  AGTTTAGCTTTTGTGCTTTTATTTACAGTGCACGAGTCCGGATCCGGTAGTCAAAATCT 724
QY      595  aagaactgcctctcaagtagagtagtgccttactctgagctcgtacaggaagtgttcttc 654
Db      725  AAGAAGCTGCTCTCTAGTAGAGTGTGCTTACTTCTTAGGCGCTGTACGGAAGTGTACTTC 784
QY      655  tgcctaaagaactcgcggaattc 676
Db      785  TGCTCTAAAGCTCGCGGATTC 806

RESULT      2
US-08-644-664B-6
: Sequence 6, Application US/08644664B
: Patent No. 5776746
:
: GENERAL INFORMATION:
:   APPLICANT: Denney Jr., Dan W.
:   TITLE OF INVENTION: Gene Amplification Methods
:   NUMBER OF SEQUENCES: 42
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Medlen & Carroll, LLP
:     STREET: 220 Montgomery Street, Suite 2200
:     CITY: San Francisco
:     STATE: California
:     COUNTRY: United States of America
:     ZIP: 94104
:
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: PatentIn Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/644,664B
:     FILING DATE: 01-May-1996

```

```

1 CLASSIFICATION: 435
2 ATTORNEY/AGENT INFORMATION:
3 NAME: Ingolia, Diane E.
4 REGISTRATION NUMBER: 40,027
5 REFERENCE/DOCKET NUMBER: GENIOTPE-00912
6 TELECOMMUNICATION INFORMATION:
7 TELEPHONE: (415) 705-8410
8 TELEFAX: (415) 397-8338
9 INFORMATION FOR SEQ ID NO: 6:
10 SEQUENCE CHARACTERISTICS:
11 LENGTH: 633 base pairs
12 TYPE: nucleic acid
13 STRANDEDNESS: double
14 TOPOLOGY: linear
15 MOLECULE TYPE: DNA (genomic)
16
17 US-08-644-664B-6
18
19 Query Match 15.6%; Score 400; DB 1; Length 633;
20 Best Local Similarity 96.1%; Pred No.1.7e-73;
21 Matches 487; Conservative 0; Mismatches 6; Indels 14; Gaps 7
22
23 QY 1 aggcgaagatgacaaacatgcatcctcaaatatgacagaaacatagctccggccct 60
24 131 AGGCAGAGATGCAAGAGATCATCTC-AATYATGACAC-AACATAATATGCC-GCCCT 187
25
26 QY 61 aactccgccatcccgcccttaactccgncacagltccggccatctccgcccatgac 120
27 188 AACTCCGCCCATCCGCCCTTAACTCGCCCC--AGTTCGCCCATCTTCGCCCATGGC 245
28
29 QY 121 tgacctaattttttatattatgcaga-gcgagggcgccctggcccttgagctatccag 179
30 246 TGACTAATTTTTTTATTATTATGACAGAGCCGAGGCCGCTCGGCTCTGACTATTCCAG 305
31
32 QY 180 aagtagagagagcctttttgagagcctagagcctttgcaaaagcctcctcgatcgag 239
33 306 AAGTAGAGAGAGCCTTTTTTGAGAGCCTAGGCTTTTGCAAAAAGCTCTCGA----- 358
34
35 QY 240 gctgcacatctcccttaagcgccggccgacctactctagagggcgacatcaagccggtg 299
36 359 GCTGCACATCTCCCTTACAGCGCCGCCGCCCTACTGTGAGCGGCATCCACGCGGTTG 418
37
38 QY 300 agtcgactctgcgcgcctccgcgcctgctgctcctgaactgcgtccgcgctctagta 359
39 419 AGTCGCGTTCTCCGCCCTCCGCTCTGCTTCTGACTGCTGCTCCGCCCTTAAGTA 478
40
41 QY 360 agttaaagctcaagtcgagacggagccttttcggcgctccctctgagagcctactaga 419
42 479 AGTTAAGAGCTCAGGTCGAGACCGGCGCTTTGTCGGCGCTCTCCCTTGAGAGCTACTATA 538
43
44 QY 420 ctgacggcgctctcaagcgtcttgctgacacctgtctgtctcaactcaagctcttgttca 479
45 539 CTCACCGCGCTCTCCACGCTTTGCTTGACCCCTGCTGCTCACTGCTACGCTTTGTTTC- 597
46
47 QY 480 gttctctgtctgcgcggttaagatc 506
48 598 GTTCTCTGTTGCGCCGTTACAGATC 624
49
50 RESULT 3
51 US-08-761-277A-6
52 Sequence 6, Application US/08761277A
53 Patent No. 5972334
54 GENERAL INFORMATION:
55 APPLICANT: Denney Jr., Dan W.
56 TITLE OF INVENTION: Vaccines For Treatment Of Lymphoma And
57 TITLE OF INVENTION: Leukemia
58 NUMBER OF SEQUENCES: 80
59 CORRESPONDENCE ADDRESS:
60 ADDRESSEE: Medlen & Carroll, LLP
61 STREET: 220 Montgomery Street, Suite 2200
62 CITY: San Francisco
63 STATE: California

```


APPLICATION NUMBER: US 07/938,161
FILING DATE: 30-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR92/00280
FILING DATE: 27-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 00137
FILING DATE: 08-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 03904
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/383
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-371-121-13

Query Match 7.1%; Score 181.4; DB 1; Length 422;
Best Local Similarity 85.6%; Pred. No. 9.3e-29;

Matches 237; Conservative 0; Mismatches 36; Indels 4; Gaps 3;

QY 239 ggtcgcacatctctcctcaagcgccgcgcgccttctgagcgcacatcaagcgcgtt 298
DB 5 GGTTCGATCTCTCCTTCAAGCGCCGCCGCTACTGAGCGCGCATCAAGCGCGG-T 63
QY 299 gattcgctctcgcgcctcgcgcctctgtgctcctctgaactgctgcgcgtctagt 358
DB 64 GACTCGGCTTCTCCGCTCCGCTCCGCTGTGCTCTCTGAACGCGCCGCTTAGGT 123
QY 359 aagtttaagctcaggtcgagcgagcgcccttctgcgcgcctcctctgagcctactag 418
DB 124 AGCTCTCAAG--GGAGCGGACAAAGGCCGCTCTGACCTGAGCTTAACTTACCTAG 181
QY 419 actcaagcgctctcagcgcttctgctgacacctgctgctcaactctagctctgttc 478
DB 182 ACTCAGCGGCTCTCCAGCGCTTGGCGTGGACCGCTTGTCTCACTACCTCTTTGTTTC 241
QY 479 agttctctgtctgcgcgttacaagatcaagctctg 515
DB 242 -GTTTCTGTCTTCCGCGCGTTACAACTTCAAGGTATG 277

RESULT 9

US-07-927-391-14

; Sequence 14, Application US/07927391
; Patent No. 6001649
; GENERAL INFORMATION:
; APPLICANT: CAPUT, Daniel
; APPLICANT: FERRARA, Pascual
; APPLICANT: MILOUX, Brigitte
; APPLICANT: MINTY, Adrian
; APPLICANT: VITA, Natalio
; TITLE OF INVENTION: Protein having a cytokin type
; TITLE OF INVENTION: activity, and recombinant DNA, expression vector and hosts
; TITLE OF INVENTION: for its preparation.
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: King Street Station, Suite 500, 1800 Diagonal
; CITY: ALEXANDRIA

STATE: VIRGINIA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/927,391
FILING DATE: 19920929
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 422 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
US-07-927-391-14

Query Match 7.1%; Score 181.4; DB 3; Length 422;
Best Local Similarity 85.6%; Pred. No. 9.3e-29;

Matches 237; Conservative 0; Mismatches 36; Indels 4; Gaps 3;

QY 239 ggtcgcacatctctcctcaagcgccgcgcgccttctgagcgcacatcaagcgcgtt 298
DB 5 GGTTCGATCTCTCCTTCAAGCGCCGCCGCTACTGAGCGCGCATCAAGCGCGG-T 63
QY 299 gattcgctctcgcgcctcgcgcctctgtgctcctctgaactgctgcgcgtctagt 358
DB 64 GACTCGGCTTCTCCGCTCCGCTCCGCTGTGCTCTCTGAACGCGCCGCTTAGGT 123
QY 359 aagtttaagctcaggtcgagcgagcgcccttctgcgcgcctcctctgagcctactag 418
DB 124 AGCTCTCAAG--GGAGCGGACAAAGGCCGCTCTGACCTGAGCTTAACTTACCTAG 181
QY 419 actcaagcgctctcagcgcttctgctgacacctgctgctcaactctagctctgttc 478
DB 182 ACTCAGCGGCTCTCCAGCGCTTGGCGTGGACCGCTTGTCTCACTACCTCTTTGTTTC 241
QY 479 agttctctgtctgcgcgttacaagatcaagctctg 515
DB 242 -GTTTCTGTCTTCCGCGCGTTACAACTTCAAGGTATG 277

RESULT 10

US-08-659-206A-1

; Sequence 1, Application US/08659206A
; Patent No. 5922685
; GENERAL INFORMATION:
; APPLICANT: Rakhimilevich, Alexander
; TITLE OF INVENTION: IL-12 Gene Therapy of Tumors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinkney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk


```

; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,327
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/586,201
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kilchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UFLA:062\KIT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 988..1728
; US-08-893-327-17

Query Match
Best Local Similarity 97.1%; Pred. No. 3.8e-25;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 505 tccaagctcggaaaaccagaagtaactcgtgaagtttagtcttttgcctttatttc 564
    |||
DB 796 TCGAGGAACTGAAAAACCAAGTTAACTGTAAGTTAGTCTTTTGTCTTTATTTC 855
QY 565 aggtcccgagtcgggtgtgtgtaacaaacgaagactccctccagtgagtggtcctt 624
    |||
DB 856 AGGTCCCGGATCCGGTGTGTGTCGAATCAAGAACTGCTCCTCAGTGGATGTTGCCCTT 915

QY 625 actctagagcctgtaacgaagtgtaactctgctctaaagctgcggaattcta 678
    |||
DB 916 ACTTCTAGGCCTGTACGGAAGTGTACTCTGCTCTAAAGCTGCGGAATTGTA 969

RESULT 13
US-08-893-327-19
; Sequence 19, Application US/08893327
; Patent No. 6020192
; GENERAL INFORMATION:
; APPLICANT: Zolotukhin, Sergei
; APPLICANT: Hauswirth, William W.
; APPLICANT: Muzyczka, Nicholas
; TITLE OF INVENTION: Humanized Green Fluorescent Protein
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
```

```

; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/893,327
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/586,201
; FILING DATE: 18-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kilchell, Barbara S.
; REGISTRATION NUMBER: 33,928
; REFERENCE/DOCKET NUMBER: UFLA:062\KIT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 988..1728
; US-08-893-327-19

Query Match
Best Local Similarity 97.1%; Pred. No. 3.8e-25;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 505 tccaagctcggaaaaccagaagtaactcgtgaagtttagtcttttgcctttatttc 564
    |||
DB 796 TCGAGGAACTGAAAAACCAAGTTAACTGTAAGTTAGTCTTTTGTCTTTATTTC 855
QY 565 aggtcccgagtcgggtgtgtgtaacaaacgaagactccctccagtgagtggtcctt 624
    |||
DB 856 AGGTCCCGGATCCGGTGTGTGTCGAATCAAGAACTGCTCCTCAGTGGATGTTGCCCTT 915

QY 625 actctagagcctgtaacgaagtgtaactctgctctaaagctgcggaattcta 678
    |||
DB 916 ACTTCTAGGCCTGTACGGAAGTGTACTCTGCTCTAAAGCTGCGGAATTGTA 969

RESULT 14
US-08-659-206A-4
; Sequence 4, Application US/08659206A
; Patent No. 5922685
; GENERAL INFORMATION:
; APPLICANT: Rakhmilievich, Alexander
; TITLE OF INVENTION: IL-12 Gene Therapy of Tumors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,206A
; FILING DATE:
; CLASSIFICATION: 435
```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J
: REGISTRATION NUMBER: 27386
: REFERENCE/DOCKET NUMBER: 110229.91144
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 608-251-5000
: TELEFAX: 608-251-9166
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 6295 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULE TYPE: other nucleic acid
: DESCRIPTION: /desc = "plasmid pMRG3196"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: join(955..1260, 1334..1675)
: OTHER INFORMATION: /product= "p35 gene product"
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 2377..3384
: OTHER INFORMATION: /product= "p40 gene product"
: US-08-659-206A-4

Query Match          6.5%: Score 166; DB 2; Length 6295;
Best Local Similarity 97.1%: Pred. No. 3.8e-25;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 505 tccaagcctgaaaaaccagaagaagttaactgtaagttacttcttcttatttc 564
    ||| | |||||
Db 641 TCGAGGAACTGAAAACGAGAAAGTTAACTGTAAGTTAGCTTTTGTCTTTATTTC 700

Qy 565 aggtcccgatccggtggtggtgcaatcaagaagactgctcctcagtgatgctt 624
    |||||
Db 701 AGGTCCGCGATCCGCTGTGTGCAATCAAGAAGTCTCTCTCAGTGATGTCCTTT 760

Qy 625 acttctagcctgtgcaagaagtgttacttctgctctaaagctgcgaattcta 678
    |||||
Db 761 ACTTCTAGGCCCTGTAAGGAGTGTACTTCTGCTTAAGCTGCGGAATTGTA 814

RESULT 15
US-08-836-022A-2
: Sequence 2, Application US/08836022A
: Patent No. 6001557
: GENERAL INFORMATION:
: APPLICANT: Trustees of the University of Pennsylvania
: APPLICANT: Wilson, James M.
: APPLICANT: Fisher, Krishna J.
: APPLICANT: Chen, Shu-Jen
: APPLICANT: Weltman, Matthew
: TITLE OF INVENTION: Improved Adenovirus Virus and
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Howson and Howson
: STREET: Spring House Corporate Cntr, P O Box 457
: CITY: Spring House
: STATE: Pennsylvania
: COUNTRY: USA
: ZIP: 19477
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/836,022A
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/331,381
```

```

: FILING DATE: 28-OCT-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Bak, Mary E.
: REGISTRATION NUMBER: 31,215
: REFERENCE/DOCKET NUMBER: GNVPN.008PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215-540-9200
: TELEFAX: 215-540-5818
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7852 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: CDNA
: US-08-836-022A-2

Query Match          6.5%: Score 166; DB 3; Length 7852;
Best Local Similarity 97.1%: Pred. No. 4.1e-25;
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 505 tccaagcctgaaaaaccagaagaagttaactgtaagttacttcttcttatttc 564
    ||| | |||||
Db 1584 TCGAGGAACTGAAAACGAGAAAGTTAACTGTAAGTTAGCTTTTGTCTTTATTTC 1643

Qy 565 aggtcccgatccggtggtggtgcaatcaagaagactgctcctcagtgatgctt 624
    |||||
Db 1644 AGGTCCGCGATCCGCTGTGTGCAATCAAGAAGTCTCTCTCAGTGATGTCCTTT 1703

Qy 625 acttctagcctgtgcaagaagtgttacttctgctctaaagctgcgaattcta 678
    |||||
Db 1704 ACTTCTAGGCCCTGTAAGGAGTGTACTTCTGCTTAAGCTGCGGAATTGTA 1757
```

Search completed: June 12, 2001, 11:36:57
Job time: 5840 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 09:27:17 ; Search time 3233 6 Seconds

(without alignments)
11680.253 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561

Sequence: 1 agcagcaagctgtgcaagca.....gsgmgaswmwawrammc 2561

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_higo_hum:*
20: em_higo_inv:*
21: em_higo_rod:*
22: em_hig_hum1:*
23: em_hig_hum2:*
24: em_hig_hum3:*
25: em_hig_hum4:*
26: em_hig_hum5:*
27: em_hig_hum6:*
28: em_hig_hum7:*
29: em_hig_hum8:*
30: em_hig_inv1:*
31: em_hig_inv2:*
32: em_hig_other:*
33: em_hig_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

```

```

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_rod:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_hig1:*
61: gb_hig2:*
62: gb_hig3:*
63: gb_hig4:*
64: gb_hig5:*
65: gb_hig6:*
66: gb_hig7:*
67: gb_hig8:*
68: gb_hig9:*
69: gb_hig10:*
70: gb_hig11:*
71: gb_hig12:*
72: gb_hig13:*
73: gb_hig14:*
74: gb_hig15:*
75: gb_hig16:*
76: gb_hig17:*
77: gb_hig18:*
78: gb_hig19:*
79: gb_hig20:*
80: gb_hig21:*
81: gb_hig22:*
82: gb_hig23:*
83: gb_hig24:*
84: gb_hig25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_r01:*
95: gb_r02:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1741.8	68.0	3228	9 AX050011	AX050011 Sequence
2	1659.8	64.8	2764	9 AX050010	AX050010 Sequence
3	602.4	23.5	2851	8 AF029260	AF029260 Gallus ga
4	569	22.2	3392	56 AB009864	AB009864 Expressio
5	556.2	21.7	4951	9 AR073576	AR073576 Sequence
6	554.2	21.6	184684	79 AL162732	AL162732 Homo sapi
7	552.6	21.6	215341	65 AC016904	AC016904 Homo sapi
8	534	20.9	8799	56 AF286077	AF286077 Expressio
9	478	18.7	478	9 AX050078	AX050078 Sequence
10	432.6	16.9	4614	56 AF285183	AF285183 Cloning v
11	418.8	16.4	1683	89 AK026847	AK026847 Homo sapi

Qy	774	accgagagagcttgcgaagatgtaacagttcttctcaacagtgagcaagcgtgagcttggcaaa	833
Db	61	ACGGGAGAGCTGGCAGAGATGTACAGTTTTTATCACAGGTGCAGGCCGTTGGCATTTGCCAAA	120
Qy	834	gctattgcatltgaagacgcgaagaagatgtagagcaaatatgttatgtctgcgcgaagaagccgcg	893
Db	121	GCTATTGCAATTGAAACACGACGAAAGAGATGGACGAAATATTGTTATTGTGCACAAAGACCCCC	180
Qy	894	cagcacacatccaaactcttcaagcaacactctactcgtctgcgcgaagaatltgaagcaact	953
Db	181	CAGCACATCTCAAACATCTTAGGCACAACTATCTACTGCTGTGAAGAAATTGAACACATTT	240
Qy	954	ggagaggaagccttgcacgtatattgtgtgtgtgagagatgaaacagacagatcaagtgcttca	1013
Db	241	GGAGGAAGGCCCTTGGCATGTATTTGTTGATGTGAGAGATGAACACACATCATCGTCTGCA	300
Qy	1014	gttgagaagaagccatcaagaagaatttggagaagaatltgatctgtgtaaatatgacagctgc	1073
Db	301	GTTGAGAAACCCATCAAGAAATTTGGAGGAATTTGATATCTGTTAAATTAATGCCAGTGGCC	360
Qy	1074	attagatttgcgaatacatatgtgagcaaacacttacaagaagaattgagctctatbatbaaagtg	1133
Db	361	ATTAGATTGACCAATATCATTTGGACACACCTACCCMAAGATTTGGATCTATGATGAAACCTG	420
Qy	1134	aacacccagagaccctactccttgcatactcaagaacgtaatccttacttltgaaaaagagcaaa	1193
Db	421	AACACCCAGAGACCACTACACTTGGACTTAAAGCATGATATCTTATTGAAAAGAGCCAAA	480
Qy	1194	gttgcatactctcaataatcaatccacacacgaaacctaaatccagtttgggttcaaaacag	1253
Db	481	GTTGCTATATCTCTCAATATCAGTCCAGCACAGTAACCTTAATTCAGTTTGGTTCAAAACAG	540
Qy	1254	caactgtctataccatctgctaagatgtgtaagtctatgtatgtgcttgcgaatgtgcagaa	1313

QY	1374	gctctcgatactctcggaagaccctgtatccgaagccagctgataaaagtctgatactt	1433
Db	661	GCCTATGGATATGCTGGGAGGACCTGGTATTCGAAGCCAGTGTAGAAAAGTTGATTTCA	720
QY	1434	gcgaatgcagcatattccattttccaaaagccaaaagtttctcctgcaacttctgcat	1493
Db	721	GCAGATGCACGATATTCGATTTTCGAAAAGCCAAAAGTTTACTGGCACTTTGTCA	780
QY	1494	gataaataatctctaaagaagaagaatcagaaatttgaagctttatgcgaattaaacca	1553
Db	781	GATGAAAATATCTTTAAAAGAAAGGAATTAAGAAAATTTTGACGTTTATGCAATTAACCA	840
QY	1554	gtctacctcttgcaaccagaatctctctctcagatgtaaaccagaagcttgacaagaa	1613
Db	841	GGTCATCCTTTGCACACAGATTTCTTCTTAGATGATTAACCAAGACGTTATGCAAGAA	900
QY	1614	gtggaatcaactcggctgtcttccagaatctcaagaagaagaacctgcagctgcaccca	1673
Db	901	GTGGAATCAACTGGTGGCTGTTCCAGAAATTCAAAAGAGAAACGACGTCGCAACCAAA	960
QY	1674	ccaagctcttggaagctctggaagaagaacattagaattgttaaagacctctcagtgat	1733
Db	961	CCACGTTCTTGAGCTGTGGGAAGAACATTTAGCAATGTTTAAGAGACTCTCCAGTATG	1020

Db 1021 GTCTTAAAGCCACTCAGCATCTATCTCTTTGACTCTCCGGTGAAGATGGTGGCAG 1080

Db 1081 TGGTTCCTGATCIGAAAGCAAGGCTGGGATGTGGATATGGAGACCTTCGATCAG 1140

QY 1299 ctggaatgagcagaagaatttaagaagtgaatttcagtcacatgcatctatgagccttaaaaca 1358
 Db 716 TTGGGAAGTGGAGAGAAATTTAAAGAGAAAGTGTCTGCAATGCTTTATGCGCTPAACAA 657
 QY 1359 gccatcacactgtctatgatatgctgtggaggaacctgtatcagaaagccagtgtaga 1418
 Db 656 GCCATACATACAGCTGCATGATATGCTGGAGAGATTGCGAATAGACAGACAGTGCAGA 597
 QY 1419 aaagtatcatcattgcagatgcagcatatccatttccaaaagccaaaagtttact 1478
 Db 596 AAAACTGCATCATGTGCAGATGCTGCATATTTGCAATTTTAAAGGCCAAAGGCTTACT 537
 QY 1479 ggcacattgcatctgataaataatcttaaaagaagaatagaaaattttagcgtt 1538
 Db 536 GGAAGCTTCAATTATGATGAGTCTACTGAGAGAAAGCGTTTANGCATTTTATGCTC 477
 QY 1539 tatgcatlaaaccaggttcaccccttgcaaccagatcttctttagaataccagaa 1598
 Db 476 TATACAAATTCACAGGCA-CCCTGATGCTGACTTCTTGGATGCTGAACCTG-- 418
 QY 1599 gcagttagcaagaagtgaatcaactgtgtcttccagaattccaagaagaagaactg 1658
 Db 419 -----ACATGACAGAAAGCCATGCTGCTTCACAGCATCATGAGCCAG 373
 QY 1659 cagctgcaaccaaaacacgctctcgtgagctgtggaagaacatttagattgttaagac 1718
 Db 372 GCTAATTAAGATTCTGCTCCAGCTGGCCCTGTGACAGAAATTCAGCTTATCCAGGGA 313
 QY 1719 tcttcagttgtagtctgtttaaagccactcaagaacatctatctgtttaaactcctcgt 1778
 Db 312 GAACCTAATAAAGAAATGTAATCCACAGGCGCTTTCAAGTTGAACGTGACAGT 253
 QY 1779 gaagaatgtgtgcaagctgttcttctgatactgaagaagaaggttgggaatgtcagata 1838
 Db 252 GATGAAGAGGCACTGTGATCTGATCTGAACCAACAGCAGCGGAGCGGG--TTTGG 195
 QY 1839 gagcttcgtatcagcagatgtgtgtagtgaatgaactgaactgaactgttgaataatg 1898
 Db 194 GAACCTCTGCAGCAGCTGATGATGAGCATGTCAGACNGCGACTTTGTGAAGATG 135
 QY 1899 ttctcagggaaactaaacacaatgtgcatctatctcagggaaattgaagaataaagt 1958
 Db 134 TTCCTGTAATTAACCAACCATGCGCTTCATGTCAGAAATTAAGATTAAGT 75
 QY 1959 aacatgcccctgaagcaatcaatttggagaagt-aatgatcagatgaatgcagactgtg 2017
 Db 74 AACATGCACTAGTAGTAAGCTAGAGAAAGATGCTGCACAGTTTAACGCTAAACTGTG 15
 QY 2018 a 2018
 Db 14 A 14
 RESULT 4
 AB009864 3392 bp DNA circular SYN 26-DEC-1997
 LOCUS Expression vector pME18S-FL3, complete sequence.
 DEFINITION AB009864
 VERSION AB009864.1 GI:2723416
 KEYWORDS
 SOURCE expression vectors plasmid:pME18S-FL3 DNA.
 ORGANISM
 artificial sequence.
 REFERENCE 1 (bases 1 to 3392)
 AUTHORS Maruyama,K. and Sugano,S.
 TITLE Direct Submission
 JOURNAL Submitted (16-DEC-1997) to the DDBJ/EMBL/Genbank databases. Sumio Sugano, The Institute of Medical Science, University of Tokyo, Department of Virology, 4-6-1, Shirokanedai, Minatoku, Tokyo 108, Japan (E-mail:sugano@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)
 REFERENCE 2 (bases 1 to 3392)
 AUTHORS Maruyama,K. and Sugano,S.

TITLE pME18S-FL3: a versatile expression vector
 JOURNAL Published Only in Database (1997) in press
 FEATURES Location/Qualifiers
 source 1..3392
 /organism="vectors"
 /plasmid="pME18S-FL3"
 /db_xref="taxon:29278"
 promoter 1..638
 polyA_site 1229..1427
 BASE COUNT 810 a 902 c 818 g 862 t
 ORIGIN
 Query Match 22.28; Score 569; DB 56; Length 3392;
 Best Local Similarity 96.88; Pred. No. 3,7e-72;
 Matches 666; Conservative 0; Mismatches 11; Indels 11; Gaps 8;
 QY 1 agcagaagatgcaaaacatgcataatcgaatttagtgaagaacatagtcgcggccct 60
 Db 131 AGCAGAGATGCAAAACATGCAATCTC-AATTAGTCAGC-AACCATAGTCCC-GCCCT 187
 QY 61 aactcgcccatcccgcccttaactccgncaccagttccggccattctcggcccatgyc 120
 Db 188 AACTCGCCCATCCCGCCCTTAAGTCCGCC--AGTTCGCCCAATTCGCCGCCATGGC 245
 QY 121 tgaataatttttattatatagcaga-gcgaagcgccctgcgcctcgaattatccag 179
 Db 246 TGACTAATTTTATTTATTTATGACAGAGCCGAGCGCCGCTCGACCTATTCAG 305
 QY 180 aagtagtaggaagccttttggagagccttagcctttgcaaaagcctct-cgatcgagg 238
 Db 306 AAGTAGTAGGAGGCTTTTGGAGGCTTAGGCTTTTGCAAAAAGCTCTCCGATCAGG 365
 QY 239 ggtcgcaatctctccttcaacgcgcgcgcctcaactcgtgagcgccatccacgcggtt 298
 Db 366 GGTGCGATCTCTCTTCAACGCGCCGCCCTCACTGAGGCGCCATCCACGCGCGGT 425
 QY 299 gagtcgcttctgcgcctccgcctgtgtgtcctccgaactgctgcgcgcgttaggt 358
 Db 426 GAGTCGCTTGTGCGCTCCCGCTGTGTGTCCTCGAACTGCGTCCGCGCTTAGGT 485
 QY 359 aagttaaagctcagtgtagacgcggccttgcgcgcctccttgaagcctcaactag 418
 Db 486 AAGTTAAAGCTCAGGTGAGACCGGCGCTTGTGCCGGCGCTCCCTTGAGACCTACCTAG 545
 QY 419 aactacgcgcctccacagcttgcgtgaacctgtgtcaactcagtccttgttctc 478
 Db 546 ACTCAGCGCGCTTCCACAGCTTGGCTGACCTGCTGCTCAACTCTACGTCCTTGTTC 605
 QY 479 agtttctgtctcgcgcgtttacagatcagaagctctgaaaaacaagaagttaactgtt 538
 Db 606 -GTTTCTGTTCTGCGCGCTTACAGATCAAGCTCTGAAAAACCAAGAAATTACTGCTA 664
 QY 539 agttagctcttctgtcttattcaggtcccgatccggtgtgtgtgcaatcaaga 598
 Db 665 AGTTAGCTTTTGTCTCTTTATTTACAGTCCGGATCC---GGTGGTCAAAATCAAGA 721
 QY 599 actgctcctcagtgatgttgccttacttctagccttgtagcgtgaagtgttaacttgc 658
 Db 722 ACTGCTCTCAGTGTGATGTTCCTTACTCTAGGCGCTGAGGAAGTACTTCTCT 781
 QY 659 cttaaagctgcggaattcttaactagact 686
 Db 782 CTAAAGCTGCGGAATTCCTCGACACT 809
 RESULT 5
 AR073576
 LOCUS AR073576 4951 bp DNA PAT 28-AUG-2000
 DEFINITION Sequence 1 from patent US 5952171.
 ACCESSION AR073576
 VERSION AR073576.1 GI:10000340
 KEYWORDS

[illegible]

```

ACCESSION      AL162732
VERSION        AL162732.28
KEYWORDS       HTG: HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE         ORGANISM
               Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 184684)
AUTHORS        Sehra, H.
TITLE          Direct Submission
               Submitted (06-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
               Requests: clonerequests@sanger.ac.uk
               On Feb 8, 2001 this sequence version replaced gi:12597074.
COMMENT        ----- Genome Center
               Center: Sanger Centre
               Center code: SC
               Web site: http://www.sanger.ac.uk
               Contact: humquery@sanger.ac.uk
               ----- Project Information
               Center project name: ba32m23
               ----- Summary Statistics
               Assembly program: XGAP4, version 4.5
               Sequencing vector: M13; M77815; 0% of reads
               Sequencing vector: Plasmid; L08752; 99% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Consensus quality: 184411 bases at least Q40
               Consensus quality: 184434 bases at least Q30
               Consensus quality: 184463 bases at least Q20
               Insert size: 184484; sum-of-contigs
               Insert size: 186775; 9.9% error; agarose-fp
               Quality coverage: 13.53x in Q20 bases; sum-of-contigs Quality
               coverage: 13.46x in Q20 bases; agarose-fp
               -----
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 3 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *
               * 1 45498: contig of 45498 bp in length
               * 45499 45598: gap of 100 bp
               * 45599 108237: contig of 62839 bp in length
               * 108238 108337: gap of 100 bp
               * 108338 184684: contig of 76347 bp in length.
               Location/Qualifiers
                 1. 184684
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="9"
                   /clone="RP11-32M23"
                   /clone_11b="RPC1-11.1"
                   1. 45498
                     /note="assembly_fragment:06222
                     fragment_chain:1"
                     45599 108237
                       /note="assembly_fragment:02558
                       fragment_chain:1"
                       108338 184684
                         /note="assembly_fragment:00830
                         fragment_chain:1"
BASE COUNT     52625 a 36409 c 38646 g 56804 t 200 others
ORIGIN
Query Match    21.6%; Score 554.2; DB 79; Length 184684;
Best Local Similarity 87.7%; Pred. No. 4.2e-70;
Matches 580; Conservative 45; Mismatches 8; Indels 28; Gaps 2
1900 ttccagggaactaaaccacaatgctcattcaggtgaattgaagattaaggtga 1959
|||||

```


KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
COMMENT
FEATURES

Cloning vector pSGHVO.
Cloning vector pSGHVO
artificial sequence: vectors.
1 (bases 1 to 4614)
Leahy,D.J., Dann,C.E. III, Longo,P., Perman,B. and Ranyar,K.K.
A mammalian expression vector for expression and purification of
secreted proteins for structural studies
Protein Expr. Purif. 20 (3), 500-506 (2000)
20541642
11087690
2 (bases 1 to 4614)
Leahy,D.J.
Direct Submission
Submitted (06-JUL-2000) Biophysics, Johns Hopkins University, 725
N. Wolfe St., Baltimore, MD 21205, USA
Mammalian expression vector: directs expression of hgh fusion
proteins.
Location/Qualifiers
1. 4614
/organism="Cloning vector pSGHVO"
/db_xref="taxon:148881"
join(700..709,970..1037)
/note="Region: human growth hormone precursor."
join(700..709,970..1130,1340..1459,1552..1716,1969..2139)
/codon_start=1
/product="human growth hormone/octahistidine/etch virus
protease recognition site fusion protein precursor"
/protein_id="AAG59857.1"
/db_xref="GI:12584848"
/translation="MAGSRTSLALFLCLPMLQESAPPTPLSLFDMIRAH
RLHDLAFDTQEFEEAYIPKEKYSFLNPQTSICFSISIPSPRRETOOKSNLEL
RISLDLSOMLEPVOFLRSVANSLVNAGSDSNVLDLKEEGIQTLMGRLDESPR
TGQLEKQYTSKEDTIPSENNDALILKNYGLCYCFKMDVETFLRIYQCRSEVSGSGFS
GHHHHHHDDIPSENLDFOGSSRAFRRAARGHA"
join(1038..1130,1340..1459,1552..1716,1969..2277)
/product="human growth hormone/octahistidine/etch virus
protease recognition site fusion protein"
2140..2163
/note="Region: octahistidine tag"
2164..2205
/note="Region: tobacco etch virus protease recognition
site"
misc-feature
2233..2284
/note="multiple cloning site"
complement(3506..4366)
/codon_start=1
/product="beta-lactamase"
/protein_id="AAG59856.1"
/db_xref="GI:12584847"
/translation="MSIQFVRLIPEFPAFCPLPVFAHPETLVKKAEDQLGARVY
IELDNKKILLESFPEREPFMSTFKVLLGAVLSRIIDAGEDQGRIRIHQSQNDLVE
YSPVTEKLDTGMIVRELCSAIIITMSDITANLLITIGSPRELPAFHHMGDHYRL
DRWEPELKEAIPNDERDITMPVAAATTLRIKLITGLLILASQQLIDMWEADKVA
GLRSALPAGMFLADKSGAGERSGRIGIALLGPDGKSRIVITVTSQNTMDERNQIA
ELGASLKNH"
BASE COUNT 1099 a 1267 c 1147 g 1101 t
ORIGIN

Query Match 16.9%; Score 432.6; DB 56; Length 4614;
Best Local Similarity 97.7%; Pred. No. 1e-52;
Matches 502; Conservative 0; Mismatches 5; Indels 7; Gaps 6;

1 aggcagaagatgcaagaatgcatcaatgaatgcaagaacatagtcggagccct 60
|||||
131 aggcagaagatgcaagaatgcatctc-aaatgacac-aaacatagatcc-gc 187
|||||
61 aactccgacatccgacccctaaactccgacccatccgacccatccgacccat 120
|||||
188 aactccgacatccgacccctaaactccgaccc-aaatgacac-aaacatagatcc 245
|||||

QY 121 tgactaatctttcttattatgacaga-gcgaagccgcctcggtctgactatccag 179
|||||
DB 246 TGACTAATCTTTTATTATTTATGACAGAGCCGAGCGCGCTCCGCGCTGAGCTATTTCCAG 305
|||||
QY 180 aagtagtgagagagccttttggagagccttagagcctttgaaaaaactcctgatacgaag 239
|||||
DB 306 AAGTAGTGAGAGAGCTTTTGTGAGAGCCTAGAGCTTTTGAAAAGCTCTTCATGAGAGG 365
|||||
QY 240 gctcgcatactcctctcaacgcgcgcgcgccttactgagagcgcgcacatccagcgcgttg 299
|||||
DB 366 GCTCGCATCTCTCTCTACAGCGCGCGCGCGCGCGCTACTACTGAGCGCGCATCAGCGCGGTG 425
|||||
QY 300 agtcgcgtctcgcgcctcgcgcctcgcgcctggtgctcctcctgtaactgctgcgcgtt 359
|||||
DB 426 AGTCGCGTCTCGCGCTCTCCCGCTGTGCTGCTCTGTAACGCTCCCGCTCTAGGTA 485
|||||
QY 360 agttcaaacatgaatgctgagacgcgcgccttctgagagcctcctctgagagactactaga 419
|||||
DB 486 AGTTTAAAGCTCAGGTGAGACCGGCGCTTTTCTCGGCGCTCTTGAGGCTTACTTACA 545
|||||
QY 420 ctcagccgcgtctcgcgcgccttgcctgacccgcctgctgctcaactcactgcttcttca 479
|||||
DB 546 CTCAGCGCGCTCTCCAGCGCTTGTGCTGACCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 604
|||||
QY 480 gttctctgctcgcgcgcgttcaagatccagatc 513
|||||
DB 605 GTTTCTGTGTGCGCGCTTACAGATCAAGCTC 638
|||||

RESULT 11

AK026847
LOCUS AK026847 1683 bp mRNA PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ23194 f1s, clone REC00490.
ACCESSION AK026847
VERSION AK026847.1 GI:10439802
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens primary epithelial cells of human renal proximal
tubule cDNA to mRNA, clone_11b:REC clone:REC00490.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
Fujitani,Y., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hiroe,M.,
Ohmori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibaara,T.,
Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NEDO human cDNA sequencing project
JOURNAL unpublished (2000)
REFERENCE 2 (bases 1 to 1683)

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
Shibaara,T., Tanaka,T. and Nakamura,Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano. Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry for Japan: cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction; 5'- & 3'-end one pass sequencing; Deparment of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).

FEATURES

source Location/Qualifiers
1..1683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="primary epithelial cells of human renal
proximal tubule"
/clone="REC00490"

[illegible]

RESULT	14
LOCUS	AX070432
DEFINITION	Sequence 904 from Patent WO0102568.
ACCESSION	AX070432
VERSION	AX070432.1
KEYWORDS	GI:12580217
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	1 (bases 1 to 390)
TITLE	Williams, L.T., Escobedo, J., Innis, M.A., Garcia, P.D., Klinger, J., Kassam, A., Reinhard, C., Randozzo, F., Kennedy, G.C., Pot, D., Lamson, G., Drmanac, R., Cikenjakov, R., Drmanac, S., Dickson, M., Labat, I., Leshkowitz, D., Kita, D., Garcia, V. and Strache-Crall, B
JOURNAL	Human genes and gene expression products
FEATURES	Patent: WO 0102568-A 904 11-JAN-2001;
source	CHIRON CORPORATION (US); HSEQ, INC. (US)
	Location/Qualifiers
	1..390
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
BASE COUNT	129 a 68 c 81 g 112 t
ORIGIN	

[illegible]

Db	314	AAATTTGACCTTTATGCAATTAACCAAGTCATCCTTTGCAACCAGATTTCTCTTGA	373
Oy	1586	tgataccagaagcag	1602
Db	374	TGAATPCCCGAGACAG	390

RESULT	15
AK000572	
LOCUS	AK000572 1755 bp mRNA
DEFINITION	Homo sapiens CDNA FLJ20565 f1s, clone REC00542.
ACCESSION	AK000572
VERSION	AK000572.1 GI:7020758
KEYWORDS	oligo capping; f1s (full insert sequence).
SOURCE	Homo sapiens primary epithelial cells of human renal proximal
	tubule CDNA to mRNA, clone_11b:REC clone:REC00542.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCES
AUTHORS
TITLE
JOURNAL

1 (sites)
Tanigami, A., Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K.,
Hirao, M., Ohmori, Y., Oca, T., Suzuki, Y., Obayashi, M., Nishi, T.,
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEO human cDNA sequencing project
unpublished (2000)
2 (bases 1 to 1755)
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
Direct Submission
Submitted (15-FEB-2000) to the DDBJ/EMBL/Genbank databases. Sumio

JOURNAL COMMENT

Submitted (15-FEB-2000) to the DBJ/EMBL/GenBank databases. Sunto Sugano, Institute of Medical Science, University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)

NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'-6' end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).

FEATURES	SOURCE
location/Qualifiers	1..1755
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/cell_type="Primary epithelial cells of human renal proximal tubule"	
/clone="REC00542"	
/clone_lib="REC"	
/note="cloning vector pME18SFL3"	
503..1489	

/translational="MSMLAEERRRKKOKMAVDPQNTNANWSDSKSEFOPLRLKMGWKGKGG
 LGADEOCATDHIKQVQVNNHLEGLGATINNEDNIAHODDFQOLIAELNTICIGETTTT
 SDPKKERSFSLSEKSKTSIKNRYVHMYFLTKGSDLSRKTDLDCIFGRGROSKRTPEGGA
 SPSPENENETTTTSAFTIOEYFAKRMALAKPKPPVPSGISTOYVRRGKRRKNE
 ATGDGVSEYSLQPKAKRHTGEKPEPERAEAOERAKKSAFAERLOLPGOWDSSKASAD
 AGHDVQPEEGDFLTKPKRRGKKKLLORPVEIAEDATLEETLYLVRKKKKKKSKK"

Query Match 13.7%; Score 351.4; DB 88; Length 1755;
 Best Local Similarity 98.1%; Pred. No. 4,1e-41;
 Matches 366; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 12, 2001, 11:52:03 ; Search time 3231.71 Seconds
(without alignments)
11687.084 Million cell updates/sec

Title: US-09-464-039-8
Perfect score: 2561
Sequence: 1 aggcagagaagtatgcagaagca.....gsgmgtrasmwawwrammc 2561

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1283235 seqs, 7373929652 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rtd:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rtd:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*

44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_v1:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_v11:*
59: gb_v12:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_rtd:*
95: gb_rtd2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	1305	51.0	3228	9 AX050011	AX050011 Sequence
2	1223	47.8	2764	9 AX050010	AX050010 Sequence
3	478	18.7	478	9 AX050078	AX050078 Sequence
4	418	16.3	184684	79 AL162732	AL162732 Homo sapi
5	418	16.3	215541	65 AC016904	AC016904 Homo sapi
6	348	13.6	390	10 AX070432	AX070432 Sequence
7	332	13.0	4614	56 AF285183	AF285183 Cloning v
8	305	11.9	1683	89 AK026847	AK026847 Homo sapi
9	249	9.7	3392	56 AB009864	AB009864 Expressio
10	249	9.7	4951	9 AR073576	AR073576 Sequence
11	249	9.7	8799	56 AF286077	AF286077 Expressio

12	242	9.4	506	58	AB030743	Human T-c
13	242	9.4	508	58	AF014658	Human T-c
14	242	9.4	508	58	AF014661	Human T-c
15	242	9.4	508	58	AF014663	Human T-c
16	242	9.4	522	59	HTU53074	Human T-cel
17	242	9.4	522	59	HTU53076	Human T-cel
18	242	9.4	522	59	AB036353	Human T-c
19	242	9.4	694	59	HT032557	Human T-cel
20	242	9.4	755	10	E02120	DNA sequenc
21	242	9.4	755	59	HLJTBH5A	Human T-cel
22	242	9.4	755	59	HTLVILTRH	Human T-Lym
23	242	9.4	756	59	HTLVILTR3	Human T-cel
24	242	9.4	757	59	HTY16483	Human T-cel
25	242	9.4	757	59	HTY16484	Human T-cel
26	242	9.4	757	59	HTY16485	Human T-cel
27	242	9.4	1755	88	AK000572	AK000572 Homo sapi
28	242	9.4	5227	10	E02746	DNA encodln
29	242	9.4	9036	59	HTU19949	Human T-cel
30	242	9.4	9068	59	HL1PROP	Human T-cel
31	238	9.3	780	59	HL1LTR	Human T-cel
32	230	9.0	1675	10	E05504	CDNA encodl
33	230	9.0	1866	10	E05503	CDNA encodl
34	230	9.0	1866	59	HTVPX	Human T-cel
35	224	8.7	407	58	AF228940	Human T-c
36	224	8.7	407	58	AF228943	Human T-c
37	224	8.7	441	58	AF228960	Human T-c
38	224	8.7	441	58	AF228961	Human T-c
39	224	8.7	454	58	AF228963	Human T-c
40	224	8.7	454	58	AF228964	Human T-c
41	224	8.7	471	58	AF228944	Human T-c
42	224	8.7	507	58	AF228949	Human T-c
43	224	8.7	507	58	AF228951	Human T-c
44	224	8.7	527	58	AF228965	Human T-c
45	224	8.7	539	58	AF228945	Human T-c

ALIGNMENTS

RESULT 1
AX050011 3228 bp DNA PAT 12-JAN-2001
LOCUS AX050011
DEFINITION Sequence 24 from Patent WO0071710.
ACCESSION AX050011
VERSION AX050011.1 GI:12226384
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 3228)
AUTHORS Denelle, P., Rosier-Montus, M.F., Arnould-Reguigne, I., Prades, C. and
Clepert, C.
TITLE Expression products of genes involved in diseases related to
cholesterol metabolism
JOURNAL Patent: WO 0071710-A 24 30-NOV-2000;
Aventis Pharma S.A. (FR)
FEATURES
Location/Qualifiers
source 1..3228
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 1077 a 578 c 655 g 910 t 8 others
ORIGIN
Query Match 51.0%; Score 1305; DB 9; Length 3228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1305; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 714 gctcgccgcgcgcgtgctgcgcgcacacctctctgatactacgaagtcacatgtaacac 773
|||||
Db 1 gctcgccgcgcgcgcgtgctgcgcgcacacctctctgatactacgaagtcacatgtaacac 60
|||||

OY 774 accggagagcttgagagatgacagttttatcaaggctgacagcttgacatggcaaa 833
|||||
Db ACCGGAGAGCTTGGAGAGATGACAGTTTATCAAGGCTGACAGCTGACATGGCAAA 120
OY 834 gctatgcatctgaaagcgaagagatgagagcgaatctgttttctgtgtaagaccgcc 893
|||||
Db GCTATTCATTTGAAGAGCAGCAAGAGATGAGCAATTTGTTATGCTGTGAAGACCCGCC 180
OY 894 caagcacatccaagaactctcagcacaatctatctctgctgcaagaatctgaagcagtc 953
|||||
Db CAGCACATCCAAACCTTTCAGCAGCAATCTATCTCTCTGTGAAGAAATGAAGCGTT 240
OY 954 ggaagaaagccttgccatglatgtgtgagagatgaacagcagatcagtgctga 1013
|||||
Db GGAGGAAGAGCCCTTGCCATGATGTGTGATGTGAGATGAAAGCAGATCAGTGTGCA 300
OY 1014 gtgagaaagccatcaagaatcttgagagatgtatcttctgttaataatgcaagcgc 1073
|||||
Db GTGAGAAAGCCATCAAGAAATTTGGAGAAATGTGTTGTTGAATTAATGCCAGTGC 360
OY 1074 attagttgaccaaatacatctgacacacacccaagaagatctgatactgataagctg 1133
|||||
Db ATTAGTTGACCAATACATTCATTCGACACACCAAGAGATTGGATCGATGATGAAGCTG 420
OY 1134 aacacagagcaccctacaccttgcaatctaaagatglatcccttaattgaaagagcaaa 1193
|||||
Db AACACAGAGCACCCTTCTGATCTTAAAGCATGTATTCCTTAATTTGAAAAGAGCAAA 480
OY 1194 gtgtctatctcccaatcatcagtcacacacacgaacctaaatcgaattgtgtcaaacag 1253
|||||
Db GTGTCTATCTCCCAATATCATTCACACACCACTTAATTCAGTTGCTTCANACAG 540
OY 1254 caactgctcatcacatctgctaaagatglatgtglatglatgtctgtaagtcagcagaa 1313
|||||
Db CACTGCTCATACCATCTGCTAAAGTATGTATGTATGTATGTATGTATGTATGTATGTAT 600
OY 1314 gaatttaagtggaatctgacagtcacatgcatlbagcctaaagaacacacacacagct 1373
|||||
Db GAATTTAAGTGGAATCTGACAGTCATTAATGATGCTTAAACAGCCATACACACAGCT 660
OY 1374 gctatgatatcttgaggagagacctggtatcgaagaagcgaagtcgaagaagtcatact 1433
|||||
Db GCTATGATATGCTGGAGAGACCTGATGCAAGCCAGTGAGAAAGTTGATATCAT 720
OY 1434 gcagatgcagacatcttccatcttccaaagcgaagaagcttctgacacacttgctcat 1493
|||||
Db GCAGATGCAGACATTCATTCATTCACAAAGCAAAAGTTTCTGCAACTTGTGCAT 780
OY 1494 gatgaataatctttaaagaagaagaagaatagaagaatctgacgtttatgcaatlaacca 1553
|||||
Db GATGAATAATCTTAAAGAGAGAAAGAAATGAAATTTGACGTTTATGCAATTTAAACA 840
OY 1554 gtcacatctctgacacacagatcttctcttagatgataaccagaagcagtcgaagaaa 1613
|||||
Db GTCATCTCTTGCACACAGATTTCTTATGATGAATACCCAGAGAGAGTTAGCAAGAAA 900
OY 1614 gtggaatcaactggtgctgcttccagaaatcgaagaagaagaactgcaagctgcaacaaa 1673
|||||
Db GTGGAATCAACTGCTGCTGTTCCAGAAATTCAAAGAAAGAACTGCAGCTGCAACAAA 960
OY 1674 ccaagctctgagagctggtggaagaagaacttagaatgttgaagaactctccatgtatgat 1733
|||||
Db CCAGCTTCTGAGAGCTGTGGAAGAAACATTTTGAATTTGTAAGACATCTTCACATGATGAT 1020
OY 1734 gtgttaagcactcaagaacatctactctgttgaacctcccggtggaagtgtgtgcaag 1793
|||||
Db GTTCTTAAAGCCACTCAAGCAATCTATCTTTGAATCTTCGGTGAAAGATGGTGACAG 1080
OY 1794 tggttcttgatctgaaagcgaaggtggaagtgtgcgataatgagagacctctgacag 1853
|||||
Db TGGTTCTTATGATCGAAGAAACGAAGGTTGGAATGTGCAATGTGAGAGAGCTTGTGATCAG 1140
OY 1854 gcagatgtgtgtatgtagtactgactgactgttgttaaaatglttcaagggaacta 1913
|||||

[illegible]

Db	Accession	Gene	Protein	Length	Start	Stop	Score	E-value	Organism
QY	1276	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1335	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1336	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1395	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1455	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1515	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1575	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1635	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1695	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1755	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1815	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1875	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1935	agta	agta	478	1	478	12.00	1.0e-100	human
QY	1995	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2055	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2115	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2175	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2235	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2295	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2355	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2415	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2475	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2535	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2595	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2655	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2715	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2775	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2835	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2895	agta	agta	478	1	478	12.00	1.0e-100	human
QY	2955	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3015	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3075	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3135	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3195	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3255	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3315	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3375	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3435	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3495	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3555	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3615	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3675	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3735	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3795	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3855	agta	agta	478	1	478	12.00	1.0e-100	human
QY	3915	agta	agta	478	1	478	12.00	1.0e-100	human
QY									

```
source 1. .478
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
BASE COUNT 154 a 99 c 104 g 121 t
ORIGIN

Query Match 18.7%; Score 478; DB 9; Length 478;
Best Local Similarity 100.0%; Pred. No. 1.8e-252;
Matches 478; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 796 cagttttatcacagtgcaagccggtgcaattggcaagctattgcatgaaagcagca 855
    |||||||
DB 1 CAGTTTATCACAGTGCAAGCCGTGCAATGGCAAGCTATTGCAATGCAAGACAGCAA 60
    |||||||

QY 856 aggatgagaataatggtattgtctgcaagaccgcccagcacatcacaactctag 915
    |||||||
DB 61 AGGATGAGCAAAATTTGTTATGCTGCAAGACGCCCAATCCAAACTTCTAG 120
    |||||||

QY 916 gcaaatctactgctgctgcaagaatgcaagcagttggaagaaggcctgcatgta 975
    |||||||
DB 121 GCACAAATCTATGCTGCTGCAAGAAATGAAACAGTTGGAGAAAGCCTTGCAATGTA 180
    |||||||

QY 976 ttgttgatgagagatgaaacagagctgctgctgcaagtggaagaagccatcaagaat 1035
    |||||||
DB 181 TTGTGATGAGAGATGAACAGACAGTCAGTCTGCAGTGAGAAAGCCATCAAGAAAT 240
    |||||||

QY 1036 ttgaggaatctgatatctgttaataatccagtgccattgattgacacatacttg 1095
    |||||||
DB 241 TTGAGAGAAATGTAATTCGTGTAATAATGCCAGTGCATTAGTTGACAAATACATTGG 300
    |||||||

QY 1096 acacacctaccaaagatgagctgcatgtaagcaagtgaacaccagaagcactacctg 1155
    |||||||
DB 301 ACACACCTACCAAGAGATGTGATCTGATGTAACGGAACACGACGACCTACCTTG 360
    |||||||

QY 1156 catctaagaatattcctcatttgaaagaaggaagtgtcatatcctaatactca 1215
    |||||||
DB 361 CATCTAAGATGATTCCTTATTGAAAAGCAAGATTGCTCATATCTCAATATCA 420
    |||||||

QY 1216 gtccacacctgaacctaaatcagctgtgttcaaacagcactgctctaaacattgc 1273
    |||||||
DB 421 GTCACACACGTAACCTAAATCAAGTTGGTTCAAAAGCACTGTCTTATACCAATTGC 478
    |||||||

RESULT 4
AL162732 184684 bp DNA HTG 07-FEB-2001
LOCUS Homo sapiens chromosome 9 clone RP11-32M23, *** SEQUENCING IN
DEFINITION PROGRESS ***, 3 unordered pieces.
ACCESSION AL162732.28 GI:12717974
VERSION AL162732
KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
          1 (bases 1 to 184684)
REFERENCE 1
AUTHORS Sehra,H.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          requests: clonerequest@sanger.ac.uk
          On Feb 8, 2001 this sequence version replaced g1:12597074.
          ----- Genome Center
          Center: Sanger Centre
          Center code: SC
          Web site: http://www.sanger.ac.uk
          Contact: humquery@sanger.ac.uk
          ----- Project Information
          Center project name: ba32m23
          ----- Summary Statistics
          Assembly program: XGAP4; version 4.5
          Sequencing vector: M13; M7815; 0% of reads
```

```
Sequencing vector: plasmid; L08752; 98% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 18441 bases at least Q40
Consensus quality: 18443 bases at least Q30
Consensus quality: 184463 bases at least Q20
Insert size: 184484; sum-of-ctnigs
Insert size: 186775; 9.9% error; agarose-fp
Quality coverage: 13.53x in Q20 bases; sum-of-ctnigs Quality
coverage: 13.46x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 45498: contig of 45498 bp in length
* 45499 45598: gap of 100 bp
* 45599 108237: contig of 62639 bp in length
* 108238 108337: gap of 100 bp
* 108338 184684: contig of 76347 bp in length.
*
FEATURES
    source
        1. .184684
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="9"
        /clone="RP11-32M23"
        /clone_1lb="RPCT-11.1"
        1. .45498
        /note="assembly-fragment:06222"
        fragment_chain:1"
        45599. 108237
        /note="assembly-fragment:02558"
        fragment_chain:1"
        108338. 184684
        /note="assembly-fragment:00830"
        fragment_chain:1"
BASE COUNT 52625 a 36409 c 38646 g 56804 t 200 others
ORIGIN

Query Match 16.3%; Score 418; DB 79; Length 184684;
Best Local Similarity 100.0%; Pred. No. 5.5e-219;
Matches 418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2045 aggaataataaaaaaagaagtcagctgctgctcaaaaaagtaaaaaagctcaagct 2104
    |||||||
DB 91849 AGGAATAATATAAAAAAAGTCAGCTGCTATGCTCAAAAAGTAATAAAAGCTCAACAGTT 91908
    |||||||

QY 2105 aaaaactaatgttcttcttctcctgtaataataagataatgcaagcttctcgtga 2164
    |||||||
DB 91909 AAATCTAATGTTGTTGTTCTTCCTGCTATTAATTAAGATATGACGTTGTTGCGA 91968
    |||||||

QY 2165 aaagatagaattgctcctcaagaagcttgaattgaataaagtgaagcgaatcaaa 2224
    |||||||
DB 91969 AAAGATAGAAATTTGCTCTAAGAGACTTGAATTAATTAATGCAAGCTAATCAAA 92028
    |||||||

QY 2225 cataagctcatcaagtgagatgagacagctgctgcttataatccaagagtttaa 2284
    |||||||
DB 92029 CATTAAGCTTATTAAGTGGATTTCTAAGACAGTCTGTGTTTATTTTCAAGGTTTAA 92088
    |||||||

QY 2285 cccttgagccttaccatcctacactgctcttctccaagaagaatatttggcgagaca 2344
    |||||||
DB 92089 CCCTTGAGCGCTTAACATCTCATTCACAGTCTTCTCCAAAGAAAAGTATTGGCGGACAA 92148
    |||||||

QY 2345 gtcaagatcaagcagtaaatagctcttcaaatctctctgttcaatgtaaaatgaagctag 2404
    |||||||
DB 92149 GTCAGATCAACAGCAATAAATTAAGTCTTTTCAAAATCTTCTGATGATGAATAACAGCTAG 92208
    |||||||

QY 2405 tctgtttaaatttttagtttggttggttatacttaagaaatcttaagatgctttt 2462
    |||||||
```


Db 4369 CCCTTTGAGCCTTACATCTCATTCACGTCTTTCTCCAGAAAAGTATTTTGGCGGACACA 4910
Qy 2345 gtccagatcaagcaagtaaatagctcttcaaatctcttcgtcaatgcaagagctag 2404
Db 4909 GTCCGATCAACGACTAAATAGCTTTTCAAACTCTTGTCTCATGTAATGAAGCTAG 4850
Qy 2405 tctgttttaaaatttttagtttgattgattatataatgaagaatcctaatgatttt 2462
Db 4849 TCTGTTTTAAATTTTATGTTTGTGATTTGATTAATGAATAATCTTAATCATGTTT 4792

RESULT 6
AX070432 390 bp DNA PAT 25-JAN-2001
LOCUS AX070432
DEFINITION Sequence 904 from patent WO0102568.
ACCESSION AX070432
VERSION AX070432.1 GI:12580217
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 390)
AUTHORS Williams,L.T., Escobedo,J., Innis,M.A., Garcia,P.D., Klinger,J.,
Kassam,A., Reinhard,C., Randazzo,F., Kennedy,G.C., Pot,D.,
Lamson,G., Drmanac,R., Ckenjakov,R., Drmanac,S., Dickson,M.,
Labat,I., Leshkowitz,D., Kita,D., Garcia,V. and Strache-Crain,B.
Human genes and gene expression products
Patent: WO 0102568-A 904 11-JAN-2001;
CHIRON CORPORATION (US); HYSEQ, INC. (US)

FEATURES
Source location/Qualifiers
1..390
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 129 a 68 c 81 g 112 t
ORIGIN

Query Match 13.6%; Score 348; DB 10; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.4e-180;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1255 acgtgttatccattgctgaagatggtatgtatgtatgtcttggaatgcaagaag 1314
Db 43 ACgtgttatccattgctgaagatggtatgtatgtatgtcttggaatgcaagaag 102
Qy 1315 aatttaaggatgaatgctgaatgcatatgagcctaaacagcatcacactgctg 1374
Db 103 AATTTAAGGAGAAATGTCAGTCAATGATGCGCTAAACAGCATACACACTGCTG 162
Qy 1375 ctatgataatgctggagagacctggtatcgaaagcagtgtagaaagttgatactatg 1434
Db 163 CTATGATATGCTGGGAGGAGCTGGTATCGAAACGAGTGAAGAGTTGATTCATTTG 222
Qy 1435 cagaatgagcatatctcatcttccaaagccaaagtttactgagcaacttgcattg 1494
Db 223 CAGATGAGCATATTTCCATTTTCCAAAAGCCAAAAGCTTTACTGCAACTTTGTCAATG 282
Qy 1495 atgaaaaatattcctaaagaagaagaaatagaanaatttgcattatgcaatlaaacag 1554
Db 283 ATGAAAATATCTTAAAGAAAGAAAGATAGAAATTTTGACGTTATGCAATTAACACAG 342
Qy 1555 gtcatccttgcagcagattctctcttagatgaatacccaagagcag 1602
Db 343 GTCATCTTTCACACAGATTTCTTCTAGATGAATACCCAGAGCAG 390

RESULT 7
AF285183 4614 bp DNA circular SYN 29-JAN-2001
LOCUS AF285183
DEFINITION Cloning vector pSGHV0, complete sequence.
ACCESSION AF285183
VERSION AF285183.1 GI:12584846

KEYWORDS
SOURCE Cloning vector pSGHV0.
ORGANISM Cloning vector pSGHV0.
REFERENCE 1 (bases 1 to 4614)
AUTHORS Leahy,D.J., Dann,C.E., III, Longo,P., Perman,B. and Ramyar,K.X.
TITLE A mammalian expression vector for expression and purification of
secreted proteins for structural studies
JOURNAL Protein Expr. Purif. 20 (3), 500-506 (2000)
MEDLINE 20541642
PUBMED 11087690
REFERENCE 2 (bases 1 to 4614)
AUTHORS Leahy,D.J.
TITLE Direct Submission
JOURNAL Submitted (06-JUL-2000) Biophysics, Johns Hopkins University, 725
N. Wolfe St., Baltimore, MD 21205, USA
COMMENT mammalian expression vector; directs expression of hgh fusion
proteins.

FEATURES
Source location/Qualifiers
1..4614
/organism="Cloning vector pSGHV0"
/db_xref="taxon:148881"
sig_peptide join(700..709,970..1037)
misc_feature join(700..709,970..1130,1340..1459,1552..1716,1969..2139)
CDS /note="Region: human growth hormone precursor"
join(700..709,970..1130,1340..1459,1552..1716,1969..2280)
/codon_start=1
/product="human growth hormone/octahistidine/etch virus
protease recognition site fusion protein precursor"
/protein_id="AAG59857.1"
/db_xref="GI:12584848"
/translation="MATGSRKTLLEAFGLCLPWLQEGSAFPIPLSRFLDNALRAH
RLHQLAFDYOEFEKAYIPKQKYSFLONFQSLCSSESIPPSNREKQKSNLELL
RISLLILQSWLEPVOFLRSYFANSVLVYASDSNVYLLDLEGIQTLNGRLEDGSPR
TGOIFKQYKSFEDTNSHNDALCKNGLYLCEFKDKVETFLRIYQCSVEGSCFS
GHHHHHHHDYDIPSEENLYFOSSRAFRPAARGA"
join(1038..1130,1340..1459,1552..1716,1969..2277)
/product="human growth hormone/octahistidine/etch virus
protease recognition site fusion protein"
2140..2163
/note="Region: octahistidine tag"
2164..2205
/note="Region: tobacco etch virus protease recognition
site"
2233..2284
/note="multiple cloning site"
complement(3506..4366)
/codon_start=1
/product="beta-lactamase"
/protein_id="AAG59856.1"
/db_xref="GI:12584847"
/translation="MSIQHRRVALIPFNAFCCLPFAHPETLVKVADEADQLGARVY
IELDLNSGKILIESFREFRPMKSTFVLLCGAVLRIDAGQQLRRIHYSNDLVE
YSPETKHLIDGATVRELCSAATITMSDNTFANLLLTIGAPKELTAFLEHMGHVRLL
DRWEPELNEATIPNDERDTPVAMATTLRLKLLTGLLTLLASRQLIDMEADKVAAPL
LRSLAPAGRIAPLAKSGAGKSGKGIITAAALPDKPRIVYITTTGQATMDENRRIA
EIASLTKRW"

BASE COUNT 1099 a 1267 c 1147 g 1101 t
ORIGIN

Query Match 13.0%; Score 332; DB 56; Length 4614;
Best Local Similarity 100.0%; Pred. No. 1.3e-171;
Matches 332; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 147 gccgagccgcctgagcctgagcctatccagaagatgaggaagcctttttgagggc 206
Db 273 GCCGAGGCCGCTCGGCTCGAGCTATTCAGAAAGTAGAGAGCGCTTTTGGAGCG 332
Qy 207 cttagcctttgcaaaaagctcctcgatcgaagagggcgcgcattcctcctcaagcgccgc 266
Db 333 CTAGCTTTTGCAAAAGCTCCTCGATCGAGAGGCGCTGCACTCTCTTACGCGCGCCG 392


```

OY 267 cgcctactgagcgccgcatccacgacggtgtagctgcttcgacgctcccgccgtgt 326
|||||
Db 393 CGCCCTACTAGAGCGCGCATCCACGCCGTTGAGTGCCTTCGCCCTCCGCCCTGCT 452
|||||
OY 327 ggtgctctcgaactgctgcgcgcgtctaggtaagttaaagctcaggtcagacggcgc 386
|||||
Db 453 GGTGCTCTCAGTAACTGGCTCCGCCGTCTAGTAGTTAAAGCTCAGTGCAGACCGGCG 512
|||||
OY 387 ctgttcgagcgctccctctgtagcctactagactcagcgctctccagcttgcctg 446
|||||
Db 513 CTTTGTCCGGGGCTCCCTTGGAGCTTACTGACTCAGCCGCTCTCCACGCTTGGCTG 572
|||||
OY 447 accgtctgtcgaactcactgcttctgttc 478
|||||
Db 573 ACCCTGCTTCTCACTCTACGCTTGTGTTTC 604
|||||

RESULT 8
AK026847 1683 bp mRNA PRI 29-SEP-2000
LOCUS AK026847 Homo sapiens cDNA: FLJ23194 fis, clone REC00490.
DEFINITION AK026847.1 GI:10439802
ACCESSION AK026847
VERSION AK026847.1
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens primary epithelial cells of human renal proximal
tubule cDNA to mRNA, clone_11b:REC clone:REC00490.
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (sites)
Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,
Fujiwara,T., Ono,T., Yamada,K., Fujii,T., Ozaki,K., Hiro,M.,
Ohnori,Y., Ota,T., Suzuki,Y., Odayashi,M., Nishi,T., Shibahara,T.,
Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.
NEDO human cDNA sequencing project
Unpublished (2000)
2 (bases 1 to 1683)
Sugano,S., Suzuki,Y., Ota,T., Odayashi,M., Nishi,T., Isogai,T.,
Shibahara,T., Tanaka,T. and Nakamura,Y.
Direct Submission
Submitted (29-AUG-2000) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, Institute of Medical Science, University of Tokyo,
Laboratory of Genome Structure Analysis, Human Genome Center;
Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan
(E-mail:cdna1@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing. Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1..1683
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="primary epithelial cells of human renal
proximal tubule"
/clone="REC00490"
/clone_11b="REC"
/note="cloning vector PME18SFL3"
BASE COUNT 506 a 333 c 303 g 541 t
ORIGIN
Query Match 11.9%; Score 305; DB 89; Length 1683;
Best Local Similarity 99.8%; Pred. No. 1e-156;
Matches 425; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
OY 251 tctcttcagcgccgacgctactgagcgccatccacgacggtgtagctgcttct 310
|||||

```

```

Db 7 TCCTTCAGCGCGCCGCCGCTTACTGAGGCGCATCCACGCCGCTTGAAGTCCGCTTCT 66
|||||
OY 311 ggcgcctccgcgcgtgtgtgctcctcgaactgagtcgcgcgtctcaggttaagct 370
|||||
Db 67 GCCGCTTCCCGCTCTGTGTGCTCTCTGAACTGCGTCCCGCTCTAGGTAACT 126
|||||
OY 371 caggtcgagcagcgcccttctcgagcgtcctcttgagacctactaactcagcgct 430
|||||
Db 127 CAGGTTCGAGACCGGGCTTTGTGTCCGGCTCTCTTGGAGCTTACTTACCTAGCAGCGGCT 186
|||||
OY 431 ctccagcgtcttgctcgaactcgtctgctcaactcactgcttctgttcaagtttctgctc 490
|||||
Db 187 CTCACGCTTTGGCTCGTACCGCTGCTTGCCTCACTACGCTTTGTTTC-GTTTCTGTTC 245
|||||
OY 491 tgcgcgcttacaagatccaagctctgaaacaccagaagtttaactgtagtttagcttc 550
|||||
Db 246 TGCGCGCTTACAGATCCCAAGCTCTGAAAAACAGAAAGTAACTGTAAGTTAGTCTTT 305
|||||
OY 551 ttgtcttattatcaggtcccgagatccggtgtgtgtgtgtaaatcaagaagctctcag 610
|||||
Db 306 TTGTCTTTTATTATTCAGTCCGCGATCCGCTGTGTGCAAAATCAAGACTGCTCTAG 365
|||||
OY 611 tggatgttgccttactctcctcgaactgagcgtgacggaagtgtactctgctcctaagctgcg 670
|||||
Db 366 TGGATGTTCCTTACTTCTTCTAGCGCTGTACGGAAGTGTACTTCTGCTCTAAAGCTGCG 425
|||||
OY 671 gaattc 676
|||||
Db 426 GAATTC 431
|||||

RESULT 9
AB009864 3392 bp DNA circular SYN
LOCUS AB009864 Expression vector PME18S-FL3, complete sequence
DEFINITION AB009864
ACCESSION AB009864
VERSION AB009864.1 GI:2723416
KEYWORDS expression vectors plasmid:pME18S-FL3 DNA.
SOURCE Plasmid vectors
ARTISAL sequence.
REFERENCE 1 (bases 1 to 3392)
Matsuyama,K. and Sugano,S.
Direct Submission
Submitted (16-DEC-1997) to the DDBJ/EMBL/Genbank databases. Sumio
Sugano, The Institute of Medical Science, University of Tokyo,
Department of Virology; 4-6-1, Shirokane-dai, Minato-ku, Tokyo 108,
Japan (E-mail:sugano@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
2 (bases 1 to 3392)
Matsuyama,K. and Sugano,S.
PME18S-FL3: a versatile expression vector
Published Only in Database (1997) In press
FEATURES
source
1..3392
/organism="vectors"
/plasmid="PME18S-FL3"
/db_xref="taxon:29278"
promoter
1..638
polya_site 1229..1427
BASE COUNT 810 a 902 c 818 g 862 t
ORIGIN
Query Match 9.7%; Score 249; DB 56; Length 3392;
Best Local Similarity 100.0%; Pred. No. 1e-125;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 230 cgatcgagggcgtcgatctctcctcgaagcgccgcgcctcactgagcgccatcc 289
|||||
Db 357 CGATCGAGGGGCTCGCATCTCTCTTCAACGCCGCCGCCCTTACTGATGAGCCGCATCC 416
|||||
OY 290 acgcggtgtgagctgcgtctcgcgcctccgcgcctgtgtgctcctcgaactgtcgcg 349
|||||

```

26-DEC-1997

```

Db 417 ACCCGGTTGAGCGCGTCTTCGCCGCTCCGCCCTCGGCGCTCTGAGACTCGTCGCC 476
OY 350 cgtctagtaagttaaaagctcagatcagacgagccgttgcgcgagcttcggag 409
Db 477 CGCTGAGTAAGTTTAAAGCTCAGGTCGAGACCGGCGCTTTGTCGGCGCTCCCTTGAG 536
OY 410 cctactagactcagcgcgcctccacgccttgcctcgaacctgtgtcactaactaagt 469
Db 537 CCTACCTAGACTCAGCGCGCTCTCAGAGCTTTCGCTGACCTGCTGCTCACTACGT 596
OY 470 cttgtttc 478
Db 597 CTTGTGTTTC 605

RESULT 10
AR073576 LOCUS AR073576 4951 bp DNA PAT 28-AUG-2000
DEFINITION Sequence 1 from patent US 5952171.
ACCESSION AR073576
VERSION AR073576.1 GI:10000340
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 4951)
AUTHORS McCarthy,S,Anthony, Gearing,D,Paul and Levinson,D,Adam.
TITLE Method for identifying genes encoding secreted or
membrane-associated proteins
JOURNAL Patent: US 5952171-A 1 14-SEP-1999;
FEATURES
source location/Qualifiers
1..4951
/organism="unknown"
BASE COUNT 1152 a 1410 c 1302 g 1087 t
ORIGIN

Query Match 9.7%; Score 249; DB 9; Length 4951;
Best Local Similarity 100.0%; Pred. No. 1.1e-125;
Matches 249; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 230 ccatgagggggtcgcgcatctctcctcagcgcgcgcgcacactgagcgccatcc 289
Db 357 CGATTCAGAGGGGCTCGCATCTCTCTCAGCGCGCGCGCGCTACCTGAGCGCGCATTC 416
OY 290 acgcgcgttgcagtcgctctgcgcgcctccgcgccttgcctcctgaactgcgcgc 349
Db 417 AGCGCGTTGAGTCGCGCTCTGCGCGCTCCGCGCTGAGTGCCTCTGAACTGCTCCG 476
OY 350 cgtctagtaagttaaaagctcagatcagacgagccgttgcgcgagcttcgcctggag 409
Db 477 CGCTGAGTAAGTTTAAAGCTCAGGTCGAGACCGGCGCTTTGTCGGCGCTCCCTTGAG 536
OY 410 cctactagactcagcgcgcctccacgccttgcctcgaacctgtgtcactaactaagt 469
Db 537 CCTACCTAGACTCAGCGCGCTCTCAGAGCTTTCGCTGACCTGCTGCTCACTACGT 596
OY 470 cttgtttc 478
Db 597 CTTGTGTTTC 605

RESULT 11
AF286077 LOCUS AF286077 8799 bp DNA circular SYN 27-DEC-2000
DEFINITION Expression vector AF286077, complete sequence.
ACCESSION AF286077
VERSION AF286077.1 GI:12034712
KEYWORDS
SOURCE Expression vector AF286077.
ORGANISM Expression vector AF286077.
artificial sequence: vectors.

```

```

REFERENCE 1 (bases 1 to 8799)
AUTHORS Wang,R.F. and Mullins,J.I.
TITLE Mammalian cell/vaccinia virus expression vectors with increased
stability of retroviral sequences in Escherichia coli: production
of feline immunodeficiency virus envelope protein
JOURNAL Gene 153 (2), 197-202 (1995)
MEDLINE 95180718
PUBMED 7875588
REFERENCE 2 (bases 1 to 8799)
AUTHORS Mullins,J.I., Hanley,T., Katsack,B.F.C. and Rowen,L.
TITLE Direct Submission
JOURNAL Submitted (11-VU-2000) Microbiology, University of Washington, HSB
K-455, Seattle, Washington 98195, USA
FEATURES
source location/Qualifiers
1..8799
/organism="Expression vector AF286077"
/db_xref="taxon:146704"
join(1..500,8708..8799)
/note="vaccinia thymidine kinase gene cluster 5' end"
46..500
/gene="thymidine kinase"
/note="derived from vaccinia; 5' end"
/pseudo
502..864
/note="derived from SV40"
872..1306
/note="SR-alpha Hybrid Promoter; formed by HTLV LTR and
SV40 promoter"
1344..1892
/note="derived from EMV"
1849..1857
/note="Oligopyrimidine tract"
1917..1934
/note="polyhistidine tract"
1960..2007
/note="similar to tpa 3' end"
2018..2288
/note="derived from bovine growth hormone"
2301..2331
/note="synthetic"
2339..2465
/note="derived from 77"
2474..2747
/note="derived from P7.5 vaccinia"
2777..3143
/note="murine beta-globin transcriptional regulation unit"
3189..3851
/note="DHFR"
3888..4141
/note="derived from bovine growth hormone"
4142..4507
/note="murine beta-globin transcriptional regulation unit"
4527..5492
/gene="neomycin resistance"
5496..5632
/note="derived from SV40 hormone"
5638..5911
/gene="thymidine kinase"
/note="derived from vaccinia; 3' end"
/pseudo
5638..6265
/note="vaccinia thymidine kinase gene cluster 3' end"
6320..6449
/note="similar to pUC19"
6453..7047
/note="p15A intermediate copy number"
7050..8192
/note="similar to pUC19"
complement(7192..8052)
/codon_start=1
/product="beta lactamase"
/protein_id="AAC45964.1"
/db_xref="GI:12034713"

```

							/translation="MSTIONHEFVALIPEFAAFNCIPVFAPHPELVIVKXAEQOLGARVGE IELDNSKLIESREPEREPRPMSTEFYVLUCGAVLSKDIOQEDELGRRIHYSQNDLVE YSPTEKHLDIGMTKVRELCSAITMSNTANLLTTTGGKEJTLALHNMGDVTVLE DRMPDELTNEAIIPNDBRDTITPMAVAAT74RLKTIELTELTLASROOLDIMMEADKVAQPD LRSLAPAGMFIAIDRSKGAGERSGRGIIALAGDPGRSRILVIYTTSOATEDMRNRQLA EGISLIKHM" 8636
misc_feature							8210 - 8636
BASE COUNT	2343 a	2028 c	2168 g		2260 t		
ORIGIN							
Query Match		9.7%	Score 249;	DB 56;	Length 8799;		
Best Local Similarity		100.0%;	Pred. No. 1.2e-125;				
Matches 249:	Conservative	0;	Mismatches	0;	Indels	0;	Gaps 0;
Oy	230	cgatcgagggggtcgcatactctcccttaacgcgcgccgcgcacctacttaggcggcatcc	289				
Dd	865	CGATGAGGGGTCCGCATCTCTCCTTCACAGGCCGCCGCCCTACTTAGGCGGCATCC	924				
Oy	230	acgcgcgttgaatgcgcgttcctgcgcgcctccgcgccttgttgctcctctgaactgcgtccg	349				
Dd	925	ACGCGGTTGAATCCGCGTTCTCCGCCCTCCGCCCTGTGTGCTCTCCTAGAATGCATCCGC	984				
Oy	350	cgctcagttaagattaaagctcaaggctagaagaccgggaccttgtccgcgcgtcccttgag	409				
Dd	985	CCTCAAGGTAAATTTPAACGTCAAGTCAAGACC GGCCCTTGTCGCCGCCCTCCCTTGAG	1044				
Oy	410	cctactagaccagcgcgcctcccaagccttgcctgaaccttgctgctcaactacgt	469				
Dd	1045	CCTACTAGACCTCACCCGGCTCTCCACGCTTGCTGCTGACCTCGCTTGCTCAACTTAAGT	1104				
Oy	470	cttgcttc 478					
Dd	1105	CTTGTCTTC 1113					
RESULT 12							
AB030743							
LOCUS	AB030743	506 bp	DNA	VRL	04-AUG-2000		
DEFINITION	Human T-cell lymphotropic virus type 1 proviral LTR sequence,						
accession	clone:KAG130.						
version	AB030743						
keywords	AB030743.1 GI:9711249						
SOURCE							
ORGANISM	.						
REFERENCE	Human T-cell lymphotropic virus type 1 (isolate:Kagoshima HTLV-I)						
AUTHORS	proviral DNA, clone:KAG130.						
TITLE	Human T-cell lymphotropic virus type 1						
REFERENCE	viruses; Retroid viruses; Retroviridae; BLV-HTLV retroviruses.						
JOURNAL	I., H., Fujiyoshi, T., Lou, H., Yashiki, S., Sonoda, S., Cartier, L.,						
AUTHORS	Nunez, L., Munoz, I., Horai, S. and Tajima, K.						
TITLE	The presence of ancient HTLV-I provirus DNA in Andean mummies						
REFERENCE	Unpublished (1999)						
JOURNAL	2 (bases 1 to 506)						
AUTHORS	L.I.H., Sonoda, S. and Tajima, K.						
TITLE	Direct Submission						
REFERENCE	Submitted (21-JUL-1999) to the DDBJ/EMBL/GenBank databases.						
JOURNAL	Hongchuan Li, Kagoshima University, Department of Virology; 8-35-1, Sakuragaoka, Kagoshima-city, Kagoshima 890-8520, Japan (E-mail:lincmed5.kufm.kagoshima-u.ac.jp, Tel:81-99-275-5283, Fax:81-99-265-8164)						
FEATURES							
Source	Location/Qualifiers						
	1..506						
	/organism="Human T-cell lymphotropic virus type 1"						
	/proviral						
	/isolate="Kagoshima HTLV-I"						
	/db_xref="taxon:11908"						
	/clone="KAG130"						
	/note="seropositive"						
LTR	1..506						
BASE COUNT	97 a	173 c	121 g		115 t		
ORIGIN							

Query Match	Best Local Similarity	Score	DB	Length
9.4%;	100.0%;	6.2e-122;	BRASP08	506;
Matches 242;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
237	GGGGTCCGATCTCCCTCCCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	296		
230	GGGGTCCGATCTCCCTCCCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	289		
237	TTGAGTCGCTTCGCTCCCTCCCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	356		
230	TTGAGTCGCTTCGCTCCCTCCCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	349		
357	GTAAGTTAAAGCTCAGCTCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	416		
350	GTAAGTTAAAGCTCAGCTCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	409		
417	AGACTCAGCGCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	476		
410	AGACTCAGCGCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	469		
477	TC 478			
470	TC 471			
RESULT 13				
LOCUS	AF014658	508 bp	DNA	VRL
DEFINITION	Human T-cell lymphotropic virus type 1 isolate BRASP08, 5' long			
ACCESSION	AF014658			
VERSION	AF014658.1	GI:3327436		
KEYWORDS				
SOURCE	Human T-cell lymphotropic virus type 1.			
ORGANISM	Human T-cell lymphotropic virus type 1.			
REFERENCE	1 (bases 1 to 508)			
AUTHORS	Yamashita,M., Veronesi,R., Mennà-Barreto,M., Harrington,M.J., Jr., Sampl,R.C., Brites,C., Badaro,R., Andrade-Filho,A.S., Okura,S., Igarashi,T., Takhisa,J., Miura,T., Chamone,D., Bianchini,O., Jardim,C., Sonoda,S. and Hayami,M.			
TITLE	Molecular epidemiology of human T-cell leukemia virus type I (HTLV-I) Brazil: the predominant HTLV-I in South America differ from HTLV-I of Japan and Africa, as well as those of Japanese immigrants and their relatives in Brazil			
JOURNAL	Virology 261 (1), 59-69 (1999)			
MEDLINE	99405662			
REFERENCE	2 (bases 1 to 508)			
AUTHORS	Yamashita,M.			
TITLE	Direct Submission			
JOURNAL	Submitted (18-JUL-1997) Institute for Virus Research, Kyoto University, Shogoin-Kawahara-Machi 53, Sakyo-Ku, Kyoto 606, Japan			
FEATURES	Location/Qualifiers			
source	1..508			
LTR	/organism="Human T-cell lymphotropic virus type 1"			
BASE COUNT	1..508			
ORIGIN	98 a 173 c 123 g 114 t			
Query Match	9.4%;	Score 242;	DB 58;	Length 508;
Best Local Similarity	100.0%;	Pred. No. 6.2e-122;		
Matches 242;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
237	GGGGTCCGATCTCCCTCCCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	296		
230	GGGGTCCGATCTCCCTCCCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	291		
297	TTGAGTCGCTTCGCTCCCTCCCTCCAGCGCGCGCCCTACTAGCGGCATCCAGCGCG	356		

Db 292 TTGAGTCGGCTCTGCGGCTCCCGCTGTGGCTCTGTAACGCTCCGCGCTCTAG 351
Oy 357 gtaagtttaaacgtacgagtcgagaccgagccctttgtcggcgctcccttgagaccact 416
Db 352 GTAAGTTTAAAGCTCAGGTGAGACCGGGCTTTGTCCGGCGCTCCCTTGAGACCTACCT 411
Oy 417 agactcagcggtctccacgcttgcctgacccgtctgtcctaactcagctcttctgt 476
Db 412 AGACTCAGCGGCTCTCCACGCTTGTGCTGACCCCTCTGTCAACTCTACGCTTTGT 471
Oy 477 tc 478
Db 472 TC 473

RESULT 14

AF014661 508 bp DNA VRL 30-AUG-1999
LOCUS Human T-cell lymphotropic virus type 1 isolate BRASP22, 5' long
DEFINITION terminal repeat sequence.

AF014661
ACCESSION AF014661.1 GI:3327439

KEYWORDS Human T-cell lymphotropic virus type 1.

SOURCE Human T-cell lymphotropic virus type 1.

ORGANISM Viruses; Retroviridae; Retroviridae; BLV-HTLV retroviruses.

REFERENCE 1 (bases 1 to 508)

Yamashita, M., Veronesi, R., Menna-Barreto, M., Harrington, W.J. Jr.,

Samplio, C., Brites, C., Badaro, R., Andrade-Filho, A.S., Okhura, S.,

Igarashi, T., Takehisa, J., Mura, T., Chamone, D., Blanchini, O.,

Jardim, C., Sonoda, S. and Hayami, M.

Molecular epidemiology of human T-cell leukemia virus type I

(HTLV-I) Brazil: the predominant HTLV-1s in South America differ

from HTLV-1s of Japan and Africa, as well as those of Japanese

immigrants and their relatives in Brazil

VIROLOGY 261 (1), 59-69 (1999)

JOURNAL

MEDLINE 99405662

REFERENCE 2 (bases 1 to 508)

Yamashita, M.

AUTHORS Direct Submission

JOURNAL Submitted (18-JUN-1997) Institute for Virus Research, Kyoto

UNIVERSITY, Shogoin-Kawahara-Machi 53, Sakyo-Ku, Kyoto 606, Japan

LOCATION/Qualifiers

1. 508

/organism="Human T-cell lymphotropic virus type 1"

/isolate="BRASP22"

/db_xref="taxon:11908"

1. 508

BASE COUNT 97 a 174 c 123 g 114 t

Query Match 9.4%; Score 242; DB 58; Length 508;

Best Local Similarity 100.0%; Pred. No. 6.2e-122; Indels 0; Gaps 0;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 237 ggggctcgatctctctcaccgagccgagcccttctgagggcgagccatcagccgag 296

Oy 237 ggggctcgatctctctcaccgagccgagcccttctgagggcgagccatcagccgag 296

Db 232 GGGGCTCGATCTCTCTTCAAGCGCGCCGCGCTCTGAGCGCGCCATCAGCGCGG 291

Oy 297 ttgagtcgctctcgcgcctccgcctctgtgtgctcctcgaactcgtcgcgcgtctag 356

Db 292 TTGAGTCGGCTCTGCGGCTCCCGCTGTGGCTCTGTAACGCTCCGCGCTCTAG 351

Oy 357 gtaagtttaaacgtacgagtcgagaccgagccctttgtcggcgctcccttgagaccact 416

Db 352 GTAAGTTTAAAGCTCAGGTGAGACCGGGCTTTGTCCGGCGCTCCCTTGAGACCTACCT 411

Oy 417 agactcagcggtctccacgcttgcctgacccgtctgtcctaactcagctcttctgt 476

Db 412 AGACTCAGCGGCTCTCCACGCTTGTGCTGACCCCTCTGTCAACTCTACGCTTTGT 471

Oy 477 tc 478
Db 472 TC 473

RESULT 15

AF014663 508 bp DNA VRL 30-AUG-1999
LOCUS Human T-cell lymphotropic virus type 1 isolate BRASP31, 5' long
DEFINITION terminal repeat sequence.

AF014663
ACCESSION AF014663.1 GI:3327441

KEYWORDS Human T-cell lymphotropic virus type 1.

SOURCE Human T-cell lymphotropic virus type 1.

ORGANISM Viruses; Retroviridae; Retroviridae; BLV-HTLV retroviruses.

REFERENCE 1 (bases 1 to 508)

Yamashita, M., Veronesi, R., Menna-Barreto, M., Harrington, W.J. Jr.,

Samplio, C., Brites, C., Badaro, R., Andrade-Filho, A.S., Okhura, S.,

Igarashi, T., Takehisa, J., Mura, T., Chamone, D., Blanchini, O.,

Jardim, C., Sonoda, S. and Hayami, M.

Molecular epidemiology of human T-cell leukemia virus type I

(HTLV-I) Brazil: the predominant HTLV-1s in South America differ

from HTLV-1s of Japan and Africa, as well as those of Japanese

immigrants and their relatives in Brazil

VIROLOGY 261 (1), 59-69 (1999)

JOURNAL

MEDLINE 99405662

REFERENCE 2 (bases 1 to 508)

Yamashita, M.

AUTHORS Direct Submission

JOURNAL Submitted (18-JUN-1997) Institute for Virus Research, Kyoto

UNIVERSITY, Shogoin-Kawahara-Machi 53, Sakyo-Ku, Kyoto 606, Japan

LOCATION/Qualifiers

1. 508

/organism="Human T-cell lymphotropic virus type 1"

/isolate="BRASP31"

/db_xref="taxon:11908"

1. 508

BASE COUNT 97 a 173 c 123 g 115 t

Query Match 9.4%; Score 242; DB 58; Length 508;

Best Local Similarity 100.0%; Pred. No. 6.2e-122; Indels 0; Gaps 0;

Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 237 ggggctcgatctctctcaccgagccgagcccttctgagggcgagccatcagccgag 296

Db 232 GGGGCTCGATCTCTCTTCAAGCGCGCCGCGCTCTGAGCGCGCCATCAGCGCGG 291

Oy 297 ttgagtcgctctcgcgcctccgcctctgtgtgctcctcgaactcgtcgcgcgtctag 356

Db 292 TTGAGTCGGCTCTGCGGCTCCCGCTGTGGCTCTGTAACGCTCCGCGCTCTAG 351

Oy 357 gtaagtttaaacgtacgagtcgagaccgagccctttgtcggcgctcccttgagaccact 416

Db 352 GTAAGTTTAAAGCTCAGGTGAGACCGGGCTTTGTCCGGCGCTCCCTTGAGACCTACCT 411

Oy 417 agactcagcggtctccacgcttgcctgacccgtctgtcctaactcagctcttctgt 476

Db 412 AGACTCAGCGGCTCTCCACGCTTGTGCTGACCCCTCTGTCAACTCTACGCTTTGT 471

Oy 477 tc 478

Db 472 TC 473

Search completed: June 12, 2001, 14:20:12

Job time: 8889 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 12, 2001, 12:41:33 ; Search time 167.51 Seconds
(without alignments)
8925.222 Million cell updates/sec

Title: US-09-464-039-8

Perfect score: 2561

Sequence: 1 aggcaggaagtgtgcaagca.....gsgmggraswmwawrrammc 2561

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 678276 seqs, 291890651 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N.Geneseq_0401:*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total distribution, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	249	9.7	4951	19 V27206	Vector pTRAP3 DNA
2	242	9.4	755	10 N92604	Recombinant DNA ve
3	242	9.4	9045	17 T42902	DNA sequence which
4	242	9.4	5	N40080	Sequence complemen
5	241	9.4	633	19 V35786	Sequence of the sp
6	230	9.0	1675	14 Q50836	HTLV-1 p21X cDNA.
7	230	9.0	1866	14 Q50835	HTLV-1 tax/rex spl
8	188	7.3	633	18 T97159	SR alpha enhancer/
9	182	7.1	720	10 N90825	Human T lymphotro
10	182	7.1	720	15 Q68398	HTLV-1 LTR genomic
11	182	7.1	720	18 T47852	HTLV-I long termin

ALIGNMENTS

12	179	7.0	198	21 C08959	Human secreted pro
13	163	6.4	3796	21 A27831	Vector plasmid pCM
14	163	6.4	6253	20 X08454	AAV vector sequenc
15	163	6.4	6280	20 X08455	AAV vector sequenc
16	163	6.4	6280	20 X08456	AAV vector sequenc
17	163	6.4	6295	19 V02043	Plasmid pMRG3196 e
18	163	6.4	6981	21 Z45931	Nucleotide sequenc
19	163	6.4	7287	19 V02042	Plasmid pMRG3169 e
20	163	6.4	7380	20 X84028	MP9 promoter beta
21	163	6.4	7612	21 Z39629	DNA sequence of pl
22	163	6.4	7852	17 T27556	Shuttle vector pad
23	163	6.4	7892	13 Q30906	Shuttle vector pad
24	163	6.4	7897	17 T27555	Shuttle vector pad
25	163	6.4	8236	18 T47202	Recombinant adenov
26	163	6.4	8509	18 T59271	Plasmid pAV.CMVlac
27	163	6.4	8509	20 X33862	AV.CMVlac2 cis pla
28	163	6.4	9077	19 V09005	Vector containing
29	163	6.4	9077	19 V09006	Vector containing
30	163	6.4	9077	21 Z45251	Expression vector
31	163	6.4	9077	21 Z45253	Expression vector
32	163	6.4	10398	17 T15286	Ad.AV.CMVlac2 hybr
33	163	6.4	10398	19 V22130	cDNA sequence of p
34	163	6.4	10398	19 V22129	cDNA sequence of p
35	163	6.4	10398	19 V22139	cDNA sequence of p
36	163	6.4	12135	18 T60555	Plasmid pAdCMVgag-
37	163	6.4	19307	17 T27558	Shuttle vector pad
38	163	6.4	36538	18 T60558	Recombinant trans
39	160	6.2	5068	20 X84027	MP9 promoter GFP
40	160	6.2	304	20 V86402	EST clone AR253.
41	104	4.1	160	21 T98740	Promoter region fr
42	104	4.1	471	21 Z98741	Promoter region fr
43	103	4.0	4249	19 V63466	Plasmid pCMTMf. C
44	98	3.8	2640	13 Q26664	bDAR. Bos taurus.
45	92	3.6	1028	16 T62780	DDP drug resistant

RESULT 1	
V27206	standard; cDNA; 4951 BP.
AC V27206;	
DT 12-OCT-1998	(first entry)
DE Vector pTRAP3 DNA sequence.	
KW Secreted protein; membrane-associated protein; protein secretion;	
KW signal peptide; alkaline phosphatase; pTRAP3; vector; reporter; ss.	
OS Homo sapiens.	
FT Key	Location/Qualifiers
FT mat_peptide	1313..2782
FT	/*tag= a
FT	/product= human placental alkaline phosphatase
PN WO9822491-A1.	
PD 28-MAY-1998.	
PF 06-NOV-1997;	97WO-US20201.
PR 19-NOV-1996;	96US-0752307.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.	
PI Gearing DP, Levinson DA, McCarthy SA;	
PI WPI: 1998-312407/27.	
DR P-PDB: W55047.	


```
FT misc_feature 1..757
FT /*tag= a
FT /label= LTR
FT polyA_signal 8584..8589
FT /*tag= b
FT misc_feature 8278..9032
FT /*tag= c
XX
XX W09630522-A1.
XX
XX 03-OCT-1996.
XX
XX 19-MAR-1996; 96MO-JP00719.
XX
XX 27-APR-1995; 95JP-0104299.
XX
XX 24-MAR-1995; 95JP-0065359.
XX
XX (SHIO ) SHIONOGI & CO LTD.
XX
XX Igarashi H, Okumura K, Orita S, Saiya A, Sakaguchi G;
XX
XX WPI; 1996-45367/45.
XX
XX DNA molecule with gene expression regulation activity - for use in
XX e.g. treatment of human T-cell leukaemia and HIV, as antiviral agent
XX and for detecting cancer
XX
XX Claim 2; Page 38-43; 77pp; Japanese.
XX
XX This sequence represents a DNA molecule with gene expression regulation
XX activity. This sequence is used in a plasmid for regulation of gene
XX expression, and treatment of viral infection pref. human T-cell leukemia
XX and HIV. The plasmid also encodes a protein which is used as an antiviral
XX agent, and also in a method for detecting cancer. The DNA molecule and
XX protein have potential uses in gene therapy, and the plasmid may also
XX have potential use in the treatment of TSP.
XX
XX SQ Sequence 9045 BP; 2086 A; 3162 C; 1713 G; 2084 T; 0 other;

Query Match 9.4%; Score 242; DB 17; Length 9045;
Best Local Similarity 100.0%; Pred. No. 2.2e-102;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 ggggctgcattctctcttaagcgcgcgcgcctactcagagcgcgcacacgcgcg 296
   |||||||
DB 352 ggggctgcattctctcttaagcgcgcgcgcctactcagagcgcgcacacgcgcg 411
   |||||||
QY 297 ttgagtcgcttgcgcgcctccgcctgtggtgctcctcgaactgctccgcgcttag 356
   |||||||
DB 412 ttgagtcgcttgcgcgcctccgcctgtggtgctcctcgaactgctccgcgcttag 471
   |||||||
QY 357 gtaagtttaagcagcagcgcgcgcgcctgtctcgcgcgcctccttgagactact 416
   |||||||
DB 472 gtaagtttaagcagcagcgcgcgcgcctgtctcgcgcgcctccttgagactact 531
   |||||||
QY 417 agactcagcgcgcctcctcagcgttgcctgaacctgtctgtaacctacttggtt 476
   |||||||
DB 532 agactcagcgcgcctcctcagcgttgcctgaacctgtctgtaacctacttggtt 591
   |||||||
QY 477 tc 478
   ||
DB 592 tc 593

RESULT 4
ID N40080
XX N40080 standard: cDNA; 9047 BP.
XX
XX N40080;
XX
XX 04-FEB-1992 (first entry)
XX
```

```
DE Sequence complementary to the genome of adult T-cell leukaemia virus
DE (ATLV) ATK-1.
XX
XX Diagnosis; therapy; prevention; leukaemia; antigen; ss.
XX
XX Adult T-cell leukaemia virus.
XX
XX Key Location/Qualifiers
XX 5' UTR 1..755
XX
XX CDS /*tag= a
XX 802..2088
XX /*tag= b
XX /product= gag
XX 2497..5184
XX /*tag= c
XX /product= pol
XX 5180..6643
XX /*tag= d
XX /product= env
XX 6834..7130
XX
XX CDS /*tag= e
XX /product= pX-I
XX 7288..7548
XX /*tag= f
XX /product= pX-II
XX 7476..7609
XX
XX CDS /*tag= g
XX /product= pX-III
XX 7622..8356
XX /*tag= h
XX /product= pX-IV
XX 8278..9032
XX /*tag= i
XX 9033..9047
XX /*tag= j
XX /label= cellular
XX
XX EPI13078-A.
XX
XX 11-JUL-1984.
XX
XX 06-DEC-1983; 83EP-0112261.
XX
XX 07-DEC-1982; 82JP-0214287.
XX
XX (NICA-) JAPAN FOUND FOR CAN.
XX (GANK-) GAN KENRYURAI ZH.
XX (JURI-) JURIDICAL FOUND.
XX
XX Yoshida M, Sugano H;
XX
XX WPI; 1984-172336/28.
XX
XX Viral genomic DNA complementary to RNA of human leukaemia virus -
XX useful in recombinant DNA producing therapeutic and diagnostic
XX proteins
XX
XX Claim 4; Table 1, Page 11-15; 23pp; English.
XX
XX Diagnosis of human leukaemia and/or lymphoma and virus infection is
XX made with part or all of the recombinant DNA. Virus antigenic
XX proteins can be produced. These peptides and proteins, and
XX antibodies against them, are useful for the diagnosis, therapy and
XX prevention of human leukaemia.
XX
XX SQ Sequence 9047 BP; 2087 A; 3164 C; 1713 G; 2083 T; 0 other;

Query Match 9.4%; Score 242; DB 5; Length 9047;
Best Local Similarity 100.0%; Pred. No. 2.2e-102;
Matches 242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 ggggctgcattctctcttaagcgcgcgcgcctactcagagcgcgcacacgcgcg 296
```

```
|||||
Db 352 ggggtcgcattctctctcctcaccgcccgcgcctaccctgagcgccattccacgccc 411
Oy 297 ttgagtcgcgtctctcgcgcctccgcctcttgctctctcctgacgcgcgcgcctcag 356
Db 412 ttgagtcgcgtctctcgcgcctccgcctcttgctctctcctgacgcgcgcgcctcag 471
Oy 357 gtaagtttaagctcaggtcagacgcgggcctcttgccgcgcctccctctgagacctact 416
Db 472 gtaagtttaagctcaggtcagacgcgggcctcttgccgcgcctccctctgagacctact 531
Oy 417 agactcagcgcgcctctcaccgcttctgctgacccctgctgctcaactcagctcttct 476
Db 532 agactcagcgcgcctctcaccgcttctgctgacccctgctgctcaactcagctcttct 591
Oy 477 tc 478
Db 592 tc 593

RESULT 5
V35786
ID V35786 standard; DNA: 633 BP.
XX
AC V35786;
XX
DT 22-SEP-1998 (first entry)
XX
DE Sequence of the specification.
XX
KW Induce; apoptosis; guanine phosphoribosyltransferase; gpt; luciferase;
KW cytokine; tumour necrosis factor; interleukin-1; inhibitory effect;
KW intracellular signal transmission; ss.
XX
OS Unidentified.
XX
PN M09822578-A1.
XX
PD 28-MAY-1998.
XX
PF 12-NOV-1997; 97WO-JP04126.
XX
PR 15-NOV-1996; 96JP-0305163.
XX
PA (CYTO-) INST CYTOSIGNAL RES INC.
XX
PI Nagasawa Y, Yoshida H;
XX
DR WPI: 1998-312464/27.
XX
PT Test system for detecting intra-cellular signal transmission
PT Inhibition - using vector containing apoptosis-inhibiting or
PT reporter gene and promoter sequence, used for, e.g. screening for
PT potential anti-inflammatory agents
XX
PS Disclosure: Fig 7; 62pp; Japanese.
XX
CC The present sequence is used in the plasmid vectors of the
CC invention. The vectors which comprise a gene which can induce
CC apoptosis under specific conditions, e.g. guanine
CC phosphoribosyltransferase (gpt), or a reporter gene, e.g. luciferase,
CC where the gene is situated downstream of a promoter which responds to
CC specific extracellular stimulation such as the presence of a cytokine,
CC e.g. tumour necrosis factor (TNF) or interleukin-1. The vector may be
CC used to transform a suitable cell line, such as a cell line which does
CC not produce hypoxanthine-guanine phosphoribosyl transferase (HGPRT). The
CC transformed cells are used to test the inhibitory effect of a gene or
CC substance on intracellular signal transmission.
XX
SQ Sequence 633 BP; 114 A; 210 C; 156 G; 153 T; 0 other;
```

Query Match 9.4%; Score 241; DB 19; Length 633;

```
Best Local Similarity 100.0%; Pred. No. 7.3e-102;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 238 ggggtcgcattctctctcctcaccgcccgcgcctaccctgagcgccattccacgcccgt 297
Db 355 ggggtcgcattctctctcctcaccgcccgcgcctaccctgagcgccattccacgcccgt 414
Oy 298 tgaagtcgctctcgcgcctccgcctctgtgtgctcctcctgaaactgctcgcgcctcag 357
Db 415 tgaagtcgctctcgcgcctccgcctctgtgtgctcctcctgaaactgctcgcgcctcag 474
Oy 358 taagtttaagctcaggtcagacgcgggcctcttgccgcgcctccctctgagacctacta 417
Db 475 taagtttaagctcaggtcagacgcgggcctcttgccgcgcctccctctgagacctacta 534
Oy 418 gactcagcgcgcctctcaccgcttctgctgacccctgctgctcaactcagctcttctgt 477
Db 535 gactcagcgcgcctctcaccgcttctgctgacccctgctgctcaactcagctcttctgt 594
Oy 478 c 478
Db 595 c 595

RESULT 6
O50836
ID O50836 standard; cDNA: 1675 BP.
XX
AC O50836;
XX
DT 09-MAY-1994 (first entry)
XX
DE HTLV-1 p21X cDNA.
XX
KW HTLV-1; human T cell leukemia virus; PCR; polymerase chain reaction;
KW detection; splice; ss.
XX
OS Homo sapiens.
XX
PN JP05244999-A.
XX
PD 24-SEP-1993.
XX
PF 18-DEC-1991; 91JP-0354839.
XX
PR 18-DEC-1991; 91JP-0354839.
XX
PA (SHRO) SHIONOGI & CO LTD.
XX
PI (SHRO) SHIONOGI & CO LTD.
XX
DR WPI: 1993-338952/43.
XX
PT Oligo-nucleotide primer - for detecting mRNA of human T cell
PT leukemia virus 1 by polymerase chain reaction
XX
PS Claim 1; Fig 8; 19pp; Japanese.
XX
CC (O50835) shows cDNA prepared from RNA extracted from a HTLV-1
CC infected cell. This was amplified by PCR and was used to construct
CC p21X cDNA (O50836). Primers and probes were then manufactured
CC (O50837-43), these are useful for the detection of HTLV-1 infection.
XX
SQ Sequence 1675 BP; 349 A; 613 C; 336 G; 377 T; 0 other;
```

Query Match 9.0%; Score 230; DB 14; Length 1675;
Best Local Similarity 100.0%; Pred. No. 8.5e-97;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 237 ggggtcgcattctctctcctcaccgcccgcgcctaccctgagcgccattccacgcccgt 296
Db 1446 ggggtcgcattctctctcctcaccgcccgcgcctaccctgagcgccattccacgcccgt 1505
Oy 297 tgaagtcgctctcgcgcctccgcctctgtgtgctcctcctgaaactgctcgcgcctcag 356

```
Db 1506 ttgagtcgcttcgcgcctccgcctctggtgctctcctgaactgctccgcgctctag 1565
Qy 357 gtaagtttaagagctcaagctcagagaccggcgcttctgcgcgcctcccttgagactact 416
Db 1566 gtaagtttaagagctcaagctcagagaccggcgcttctgcgcgcctcccttgagactact 1625
Qy 417 agactcaagcgcgctctcagagcttctgctgaacctgtctgaactcta 466
Db 1626 agactcaagcgcgctctcagagcttctgctgaacctgtctgaactcta 1675

RESULT 7
ID 050835
AC 050835;
DE 09-MAY-1994 (first entry)
HTLV-1 tax/rex splice region.
KW HTLV-1; human T cell leukemia virus; PCR; polymerase chain reaction;
KW detection; splice; ss.
OS Homo sapiens.
PN JP05244999-A.
PD 24-SEP-1993.
PE 18-DEC-1991; 91JP-0354839.
PR 18-DEC-1991; 91JP-0354839.
PA (SHIO ) SHIONOGI & CO LTD.
DR WPI; 1993-338952/43.
XX
XX Oligo-nucleotide primer - for detecting mRNA of human T cell
PT leukemia virus 1 by polymerase chain reaction
XX
XX
PS Disclosure; Fig 7; 19pp; Japanese.
CC (Q50835) shows cDNA prepared from RNA extracted from a HTLV-1
CC infected cell. This was amplified by PCR and was used to construct
CC p21X cDNA (Q50836). Primers and probes were then manufactured
CC (Q50837-43), these are useful for the detection of HTLV-1 infection.
XX
SQ Sequence 1866 BP; 394 A; 678 C; 386 G; 408 T; 0 other;

Query Match 9.0%; Score 230; DB 14; Length 1866;
Best Local Similarity 100.0%; Pred. No. 8.5e-97;
Matches 230; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 237 ggggctgcgcatctctctctcaagcgccgcgcgcctactctgagggcgcatccagcgcg 296
Db 1637 ggggctgcgcatctctctctcaagcgccgcgcgcctactctgagggcgcatccagcgcg 1696
Qy 297 ttgagtcgcttcgcgcctccgcgcctggtgctctcctgaactgctccgcgctctag 356
Db 1697 ttgagtcgcttcgcgcctccgcgcctggtgctctcctgaactgctccgcgctctag 1756
Qy 357 gtaagtttaagagctcaagctcagagaccggcgcttctgcgcgcctcccttgagactact 416
Db 1757 gtaagtttaagagctcaagctcagagaccggcgcttctgcgcgcctcccttgagactact 1816
Qy 417 agactcaagcgcgctctcagagcttctgctgaacctgtctgaactcta 466
Db 1817 agactcaagcgcgctctcagagcttctgctgaacctgtctgaactcta 1866
```

```
RESULT 8
ID T97159 standard; DNA; 633 BP.
AC T97159;
DE 11-MAY-1998 (first entry)
KW SR alpha enhancer/promoter.
DE Vaccine: B-cell malignancy; lymphoma; leukaemia; tumour;
KW gene amplification; immunotherapy; therapy; SV40; promoter;
KW enhancer; ds.
OS Chimeric - Rhesus macaque polyoma virus.
OS Chimeric - Human T cell leukaemia virus type 1.
PN W09741244-A1.
PD 06-NOV-1997.
PE 25-APR-1997; 97WO-US07039.
PR 06-DEC-1996; 96US-0761277.
PR 01-MAY-1996; 96US-0644664.
XX (GENI-) GENITOPE CORP.
XX
XX Denney DW;
PI WPI; 1997-549743/50.
DR
XX
XX Multivalent vaccine to treat B cell lymphoma or leukaemia -
PT comprises at least 2 different recombinant variable regions of
PT immunoglobulin molecules derived from B cell lymphoma cells
XX
XX Example 1; Page 104; 17pp; English.
XX
XX This DNA sequence comprises the HindIII/XhoI fragment of the
CC SR alpha enhancer/promoter in plasmid pCDL-SR alpha 296. The
CC SR alpha enhancer/promoter is composed of human T cell leukaemia
CC virus 1 5' untranslated sequences and the SV40 enhancer. It is
CC reported to increase expression from the SV40 enhancer/promoter by
CC 10-fold in host cells, and is active in a broad range of cell
CC types. The SR alpha enhancer/promoter has been utilised in
CC expression vectors designed for efficient expression of genes in
CC eukaryotic cells. The invention provides vectors and improved
CC methods for the expression and co-amplification of genes encoding
CC recombinant proteins in cultured cells. The methods permit the
CC isolation of cell lines which have co-amplified input recombinant
CC sequences which encode an amplifiable marker, one or more
CC expression vectors encoding a protein of interest and optionally a
CC selectable marker. The amplified cells provide large quantities
CC of recombinant proteins suitable for immunotherapy for treatment of
CC lymphomas and leukaemias. The methods permit the production of
CC custom vaccines, including multivalent vaccines that reflect the
CC degree of somatic variation found in a patient's tumour.
XX
SQ Sequence 633 BP; 114 A; 210 C; 155 G; 154 T; 0 other;

Query Match 7.3%; Score 188; DB 18; Length 633;
Best Local Similarity 99.6%; Pred. No. 2.5e-77;
Matches 238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 240 gctgcgcatctcctctcaagcgccgcgcgcctactgagggcgcatccagcgctg 299
Db 359 gctgcgcatctcctctcaagcgccgcgcgcctactgagggcgcatccagcgctg 418
Qy 300 agtcgcttcgctcgcctccgcgcctctggtgctcctctgaactgctccgcgctctag 359
Db 419 agtcgcttcgctcgcctccgcgcctctggtgctcctctgaactgctccgcgctctag 478
```


SQ Sequence 720 BP; 138 A; 252 C; 157 G; 173 U; 0 other;

Matches 131; Conservative 51; Mismatches 0; Indels 0; Gaps 0;

Db 239 uc 240

ID T47852 standard; RNA; 720 BP.

AC T47852;

DT 21-MAY-1997 (first entry)
XX

HTLV-I long terminal repeat region oligonucleotide.

OS Human T-cell lymphotropic virus type I.

PN US5580761-A.

PD 03-DEC-1996
XY

PF 16-FEB-1988; 88US-0156188.
YY

PR	23-MAR-1994;	94US-0217210.
PR	16-FEB-1988;	88US-0156188

PR 29-JUL-1991; 9105-0739718.
XX

PA (GREN) GREATBATCH GEN-AID LTD.
XX
XX

P1 Grealbatch W, Sanford JC;
XX

WFL, 1997-0333/1/03.
XX

PT able to inhibit replication

PS Disclosure; Column 23-24; 29pp; English.

XX

CC retroviral infection.
yy

50 Sequence 720 BP; 138 A; 252 C; 157 G; 173 U; 0 other;

Matches 131; Conservative 51; Mismatches 0; Indels 0; Gaps 0;

Dd 59 uugagucgcgaccccgccnugnugncccaacugcguccgcgucuaag 118

Db 119 guaguuuaagcucagcagaccgagccuuguccgagccuaccu 178

Db 179 agacucagccgcucuccacgcgcucgaccgcucgucacacucacgucucuu 238

Db 239 uc 240

РЕСПИТ 13

C08959
ID C0895

XX
AC
C0895

DT 06-OCT-2000 (first entry)

DE Human secreted protein 5' EST, SEQ ID NO: 13034.
 VV

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation, KW therapy; chromosome mapping; ss

Homo sapiens
XX
XX
XX

XX
PN
EP1033401-A2

XX
PD 06-SEP-2000

AA 21-FEB-2000: 2000EP-0200610.

XX 26-FEB-1999: 99US-0122487.
PR

PA (GEST) GENSET.

aa Dumas Milne Edwards J, Duclert A, Giordano J;
PI

DR WPI; 2000-500381/45.

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for

PT diagnostic, forensic, gene therapy and chromosome mapping procedures -

PS Claim 1; SEQ ID 13034; 71pp + CD-ROM; English.

CC The present sequence is one of a large number of 5' ESTs derived from

DR WPI; 1999-132241/11.
DR P-PSDB; W96328.

XX Humanised green fluorescent protein - used to measure gene
PT expression and identify transformed cells
XX
PS Claim 2; Page 132-136; 152pp; English.
XX
CC Humanised green fluorescent protein (gfp) genes can be used to
CC identify transformed cells, to measure gene expression in vitro and
CC in vivo, to label specific cells in multicellular organisms (e.g. to
CC study cell lineage's), to label and locate fusion proteins, and to
CC study intracellular trafficking. Commonly used reporter genes include
CC beta-galactosidase, firefly luciferase, alkaline phosphatase,
CC chloramphenicol acetyltransferase (CAT), and beta glucuronidase
CC (GUS). However, these have limitations in their use. Frequently,
CC these reporter genes require the addition of a substrate and the
CC size of certain proteins means that the expression of reporter
CC fusion proteins can be difficult. The light stimulated GFP
CC fluorescence is species independent and does not require any
CC cofactors substrates or additional gene products from Aequorea
CC victoria an as the GFP genes have been humanised, they are
CC expressed at sufficient levels to be detectable in human cells,
CC unlike previous GFP proteins.
CC
XX
SQ Sequence 6253 BP; 1505 A; 1624 C; 1681 G; 1443 T; 0 other;
Query Match 6.4%; Score 163; DB 20; Length 6253;
Best Local Similarity 100.0%; Pred. No. 8.4e-66;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 513 ctgaaacacgaagtaactgtaagttaagttcttttgcctttattcaggtccg 572
DB 804 ctgaaacacgaagtaactgtaagttaagttcttttgcctttattcaggtccg 863
OY 573 gatccggtggtgtaacatacaagaactgctcctcagtgagttgtccttactctag 632
DB 864 gatccggtggtgtaacatacaagaactgctcctcagtgagttgtccttactctag 923
OY 633 gctctacggaagtgttactctgtctctaaagtcgcggaatt 675
DB 924 gctctacggaagtgttactctgtctctaaagtcgcggaatt 966
RESULT 15
X08455
ID X08455 standard; DNA: 6280 BP.
XX
AC X08455;
XX
DT 28-JUN-1999 (first entry)
XX
DE AAV vector sequence comprising humanised green fluorescent protein.
XX
KW Green fluorescent protein; gfp; jellyfish; Aequorea victoria;
KW humanisation; reporter gene; substrate; cofactor; beta galactosidase;
KW firefly luciferase; alkaline phosphatase;
KW chloramphenicol acetyltransferase; CAT; beta glucuronidase; GUS; ss.
XX
OS Synthetic.
XX
XX Key Location/Qualifiers
FT CDS 988..1731
FT /tag= a
FT /product= "Humanised green fluorescent protein"
XX
XX MO9903997-A1.
XX
XX 28-JAN-1999.
XX
XX 16-JUL-1998; 98WO-US14692.
XX
XX 16-JUL-1997; 97US-0893327.
XX

PA (UYFL) UNIV FLORIDA.
XX
XX Hauswirth W, Muzyczka N, Zolotukhin S;
PI
XX
DR WPI; 1999-132241/11.
XX
DR P-PSDB; W96329.
XX
PT Humanised green fluorescent protein - used to measure gene
PT expression and identify transformed cells
XX
PS Claim 3; 137-141; 152pp; English.
XX
CC Humanised green fluorescent protein (gfp) genes can be used to
CC identify transformed cells, to measure gene expression in vitro and
CC in vivo, to label specific cells in multicellular organisms (e.g. to
CC study cell lineage's), to label and locate fusion proteins, and to
CC study intracellular trafficking. Commonly used reporter genes include
CC beta-galactosidase, firefly luciferase, alkaline phosphatase,
CC chloramphenicol acetyltransferase (CAT), and beta glucuronidase
CC (GUS). However, these have limitations in their use. Frequently,
CC these reporter genes require the addition of a substrate and the
CC size of certain proteins means that the expression of reporter
CC fusion proteins can be difficult. The light stimulated GFP
CC fluorescence is species independent and does not require any
CC cofactors substrates or additional gene products from Aequorea
CC victoria an as the GFP genes have been humanised, they are
CC expressed at sufficient levels to be detectable in human cells,
CC unlike previous GFP proteins.
CC
XX
SQ Sequence 6280 BP; 1515 A; 1627 C; 1692 G; 1446 T; 0 other;
Query Match 6.4%; Score 163; DB 20; Length 6280;
Best Local Similarity 100.0%; Pred. No. 8.4e-66;
Matches 163; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 513 ctgaaacacgaagtaactgtaagttaagttcttttgcctttattcaggtccg 572
DB 804 ctgaaacacgaagtaactgtaagttaagttcttttgcctttattcaggtccg 863
OY 573 gatccggtggtgtaacatacaagaactgctcctcagtgagttgtccttactctag 632
DB 864 gatccggtggtgtaacatacaagaactgctcctcagtgagttgtccttactctag 923
OY 633 gctctacggaagtgttactctgtctctaaagtcgcggaatt 675
DB 924 gctctacggaagtgttactctgtctctaaagtcgcggaatt 966

Search completed: June 12, 2001, 14:19:58
Job time: 5905 sec

into the Not I and Hind III sites of the Laminid BA vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 150 a 64 c 89 g 172 t 7 others

ORIGIN

Query Match 12.7%; Score 326; DB 156; Length 482;
Best Local Similarity 100.0%; Pred. No. 4.2e-161;
Matches 326; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2137 ttataagataatgcagctgttctcgtggaagaatagattctcttaagaacttgaa 2196
Db 1 ttatagagatgacagctgttctcgtggaagaatagattctcttaagaacttgaa 60

Oy 2197 ttgtaattaaatggcaagctaaatcaataagcttcttaagtgatcttaagacg 2256
Db 61 ttgtaattaaatggcaagctaaatcaataagcttcttaagtgatcttaagacg 120

Oy 2257 tctgtgtttatattcaagggttaacccttgagcctcatctcatctcattctt 2316
Db 121 tctgtgtttatattcaagggttaacccttgagcctcatctcatctcattctt 180

Oy 2317 tctcaagaagaatgttttggcgagcagctcagatcaagcagataaataagcttcaa 2376
Db 181 tctcaagaagaatgttttggcgagcagctcagatcaagcagataaataagcttcaa 240

Oy 2377 atctctgtcatgtaaatgagctagctgtttaaatttttaagtttgatgtat 2436
Db 241 atctctgtcatgtaaatgagctagctgtttaaatttttaagtttgatgtat 300

Oy 2437 actaatgaataatcttaagtattt 2462
Db 301 actaatgaataatcttaagtattt 326

RESULT 14
LOCUS AA622988/c 386 bp mRNA EST 21-OCT-1997
DEFINITION np58h04.s1 NCI-CGAP Br2 Homo sapiens cDNA clone IMAGE:1130551 3'
sequence similar to WP:CI7G10.8 CE02490 ALCOHOL DEHYDROGENASE ;, mRNA

ACCESSION AA622988
VERSION AA622988.1 GI:2526864
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 386)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LMW at:
www-bio.llnl.gov/dbip/image/image.html
Insert Length: 937 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 342.
Location/Qualifiers
1..386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1130551"

FEATURES
SOURCE

/clone.lib="NCI CGAP-Br2"
/sex="female, pooled"
/tissue_type="breast"
/lab_host="DH10B"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from pooled bulk
breast tumor tissue, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified p773
vector. This library is the normalized version of
NCI CGAP-Br1.1. Library was constructed by Bento Soares
and M. Fatima Bonaldo."

BASE COUNT 99 a 92 c 58 g 137 t

ORIGIN

Query Match 12.7%; Score 325; DB 9; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.4e-160;
Matches 325; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1694 agaaacatttagaattgttaaggaactctcagatgattgttaagcaacttaagc 1753
Db 386 agaaacatttagaattgttaaggaactctcagatgattgttaagcaacttaagc 327

Oy 1754 aatctatcgtttgacctcgcgtggaagatgagtgatgtttctgattcgaag 1813
Db 326 aatctatcgtttgacctcgcgtggaagatgagtgatgtttctgattcgaag 267

Oy 1814 caagggttggaatgtcagatgagagcctctatcagcagatgttgatgagat 1873
Db 266 caagggttggaatgtcagatgagagcctctatcagcagatgttgatgagat 207

Oy 1874 gactactgtagcttctgtaaaatgtttcagggaaactaaacacaaatggcattcat 1933
Db 206 gactactgtagcttctgtaaaatgtttcagggaaactaaacacaaatggcattcat 147

Oy 1934 gtcaaggaaattgaagatgaaggaacatgagccttagaataatggagaagctat 1993
Db 146 gtcaaggaaattgaagatgaaggaacatgagccttagaataatggagaagctat 87

Oy 1994 gaatcagatgaatgacagactgtga 2018
Db 86 gaatcagatgaatgacagactgtga 62

RESULT 15
LOCUS 219446 340 bp mRNA EST 10-FEB-1993
DEFINITION HSB28F112 STRATAGENE Human skeletal muscle cDNA library, cat.
#936215. Homo sapiens cDNA clone 28F11, mRNA sequence.

ACCESSION 219446
VERSION 219446.1 GI:29287
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 340)
Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C., Devignes
M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B., Lorenzo, F.,
Mitchell, H., Mariage-Samson, R., Pletu, G., Pouliot, Y.,
Sedastiani-Kadachis, C. and Tessier, A.
IMAGE: molecular integration of the analysis of the human genome
and its expression

JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'interationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698

FEATURES
SOURCE

1..386
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1130551"

THIS PAGE BLANK (USPTO)

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☒ **SKEWED/SLANTED IMAGES**
- ☒ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

THIS PAGE BLANK (USPTO)